The Annals of Statistics
2002, Vol. 30, No. 4, 1134–1154

CANONICAL CORRELATION ANALYSIS AND REDUCED RANK REGRESSION IN AUTOREGRESSIVE MODELS

BY T. W. ANDERSON

Stanford University

When the rank of the autoregression matrix is unrestricted, the maximum likelihood estimator under normality is the least squares estimator. When the rank is restricted, the maximum likelihood estimator is composed of the eigenvectors of the effect covariance matrix in the metric of the error covariance matrix corresponding to the largest eigenvalues [Anderson, T. W. (1951). *Ann. Math. Statist.* **22** 327–351]. The asymptotic distribution of these two covariance matrices under normality is obtained and is used to derive the asymptotic distributions of the eigenvectors and eigenvalues under normality. These asymptotic distributions differ from the asymptotic distributions when the regressors are independent variables. The asymptotic distribution of the reduced rank regression is the asymptotic distribution of the least squares estimator with some restrictions; hence the covariance of the reduced rank regression is smaller than that of the least squares estimator. This result does not depend on normality.

1. Introduction. The objective of canonical correlation analysis is to discover and use linear combinations of the variables in one set that are highly correlated with linear combinations of variables in another set. The linear combinations define canonical variables, and the correlations between corresponding canonical variables are the canonical correlations. See Anderson (1984), Chapter 12, for an exposition and Anderson (1999a) for asymptotic theory in case of regression on independent variables. In time series analysis the variables in one set are measurements made in the present, and the variables in the other set are measurements made in the past. The linear combinations of the variables of the past may be employed for predicting variables of the present and future. A framework within which to develop this analysis may be provided by autoregressive moving average models (ARMA). Box and Tiao (1977) and Tiao and Tsay (1989) have explored this area. See Reinsel (1997) and Reinsel and Velu (1998) for more details.

In this paper the asymptotic distribution of the squares of the canonical correlations between the present and past is derived for (stationary) autoregressive processes. That distribution is contrasted with the asymptotic distribution of the canonical correlations for two sets of variables with a joint normal distribution, a common regression model. The comparison mirrors the comparison between

Received August 1999; revised September 2001.
AMS 2000 subject classifications. 62H10, 62M10.
Key words and phrases. Canonical correlations and vectors, eigenvalues and eigenvectors, asymptotic distributions.

scalar versions of the models; for a scalar autoregression with process correlation ρ, the asymptotic variance of the sample correlation coefficient is $1-\rho^2$, but the asymptotic variance of the sample (Pearson) correlation in a bivariate normal distribution with correlation ρ is $(1-\rho^2)^2$. Note the variance in the bivariate model is $1-\rho^2$ times the variance in the autoregressive model. Moreover, the sample canonical correlations in autoregression are asymptotically correlated while in regression they are asymptotically uncorrelated when the population canonical correlations are distinct. Thus there are differences in inference in the two types of models; these differences affect model-selection procedures.

The asymptotic distribution of the coefficients of the canonical variables is obtained for the first-order autoregressive process. The result is of interest because it shows that the asymptotic distribution of the coefficients of the canonical variables as well as the asymptotic distribution of the canonical correlations is different from those distributions for the classical regression model. Velu, Reinsel and Wichern (1986) have also derived the asymptotic distribution of these canonical correlations, but their results differ from the results in this paper. More detail is given in Section 7.

The reduced rank regression estimator of a regression matrix is composed of a small number of canonical variables [Anderson (1951)]. It is the maximum likelihood estimator in the autoregression model as well as the regression model when the unobserved random disturbances (or errors) are normally distributed [Anderson (1951), page 345]. In Section 5 its asymptotic distribution is found under the assumption that the disturbance in the autoregressive model is independent of the lagged variables. This asymptotic distribution is the same as the asymptotic distribution of the reduced rank regressor estimator of the coefficient matrix in the classical regression model as obtained by Ryan, Hubert, Carter, Sprague and Parrott (1992), Schmidli (1995), Stoica and Viberg (1996) and Reinsel and Velu (1998), under the assumption that the disturbances are normally distributed and by Anderson (1999b) under general conditions. Here it is also shown that the asymptotic distribution is valid regardless of the assumption on the disturbances (as long as their variances exist). The algebra here differs from the algebra in Anderson (1999b) because the transformation to canonical form differs in the two models.

2. The model and canonical analysis.

2.1. *The model.* In this paper the model is the autoregressive process (AR)

$$\mathbf{Y}_t = \sum_{i=1}^{m} \mathbf{B}_i \mathbf{Y}_{t-i} + \mathbf{Z}_t, \qquad t = \ldots, -1, 0, 1, \ldots, \tag{2.1}$$

where $\mathbf{Y}_t$ and $\mathbf{Z}_t$ are p-component vectors with $\mathbf{Z}_t$ unobserved and independent of $\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \ldots$ and $\mathcal{E}\mathbf{Z}_t = \mathbf{0}$, $\mathcal{E}\mathbf{Z}_t\mathbf{Z}_t' = \mathbf{\Sigma}$. We assume that the roots of

$$\left|\lambda^m \mathbf{I} - \sum_{i=1}^{m} \lambda^{m-i}\mathbf{B}_i\right| = 0 \tag{2.2}$$

are less than 1 in absolute value. Then (2.1) defines the stationary process $\{\mathbf{Y}_t\}$ with $\mathcal{E}\mathbf{Y}_t = \mathbf{0}$ and autocovariances

$$\mathbf{\Gamma}_h = \mathcal{E}\mathbf{Y}_t\mathbf{Y}_{t-h}' = \mathcal{E}\mathbf{Y}_{t+h}\mathbf{Y}_t' = \mathbf{\Gamma}_{-h}' \tag{2.3}$$

[e.g., Anderson (1971), Chapter 7]. If $\mathbf{Y}_{t-i}$ is replaced by $\mathbf{Y}_{t-i} - \boldsymbol{\mu}$, $i = 0, 1, \ldots, m$, in (2.1), then $\mathcal{E}\mathbf{Y}_t = \boldsymbol{\mu}$. If $m = 1$, $\mathbf{Y}_t = \sum_{s=0}^{\infty} \mathbf{B}_1^s\mathbf{Z}_{t-s}$ and $\mathbf{\Gamma}_h = \mathbf{B}_1^h\mathbf{\Gamma}_0$.

2.2. *Population.* In the AR(m) model the present is represented by $\mathbf{Y}_t$ and the (relevant) past by $\widetilde{\mathbf{Y}}_{t-1} = (\mathbf{Y}_{t-1}', \ldots, \mathbf{Y}_{t-m}')'$. Let

$$\mathcal{E}\mathbf{Y}_t\widetilde{\mathbf{Y}}_{t-1}' = [\mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_m] = \mathbf{\Gamma}, \qquad \mathcal{E}\widetilde{\mathbf{Y}}_{t-1}\widetilde{\mathbf{Y}}_{t-1}' = \widetilde{\mathbf{\Gamma}}. \tag{2.4}$$

Then the population canonical correlations between $\mathbf{Y}_t$ and $\widetilde{\mathbf{Y}}_{t-1}$ are the roots of

$$\begin{vmatrix} -\rho\mathbf{\Gamma}_0 & \mathbf{\Gamma} \\ \mathbf{\Gamma}' & -\rho\widetilde{\mathbf{\Gamma}} \end{vmatrix} = 0, \tag{2.5}$$

and the canonical vectors are the corresponding solutions of

$$\begin{bmatrix} -\rho\mathbf{\Gamma}_0 & \mathbf{\Gamma} \\ \mathbf{\Gamma}' & -\rho\widetilde{\mathbf{\Gamma}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\omega} \end{bmatrix} = \mathbf{0}, \qquad \boldsymbol{\alpha}'\mathbf{\Gamma}_0\boldsymbol{\alpha} = 1,\ \boldsymbol{\omega}'\widetilde{\mathbf{\Gamma}}\boldsymbol{\omega} = 1. \tag{2.6}$$

The largest root of (2.5), say ρ_1, is the first canonical correlation and the corresponding solution to (2.6) defines the first pair of canonical vectors, which are the coefficients of the first pair of canonical variables. The first canonical correlation is the maximum correlation between linear combinations of $\mathbf{Y}_t$ and $\widetilde{\mathbf{Y}}_{t-1}$.

From (2.6) we obtain

$$\mathbf{\Gamma}\widetilde{\mathbf{\Gamma}}^{-1}\mathbf{\Gamma}'\boldsymbol{\alpha} = \rho^2\mathbf{\Gamma}_0\boldsymbol{\alpha}. \tag{2.7}$$

There are p nonnegative solutions to

$$\left|\mathbf{\Gamma}\widetilde{\mathbf{\Gamma}}^{-1}\mathbf{\Gamma}' - \rho^2\mathbf{\Gamma}_0\right| = 0 \tag{2.8}$$

and p corresponding linearly independent vectors $\boldsymbol{\alpha}$ satisfying (2.6) and $\alpha_{ii} > 0$. If the root ρ^2 of (2.8) is unique, the solution $\boldsymbol{\alpha}$ of (2.6) (and $\alpha_{ii} > 0$) is uniquely determined. [Since the matrix $\mathbf{A} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_p)$ is nonsingular, the components of $\mathbf{Y}_t$ can be numbered in such a way that the ith component of $\boldsymbol{\alpha}_i$ is nonzero.] The number of positive solutions to (2.8) is equal to the rank of $\mathbf{\Gamma}$ and indicates the degree of dependence between $\mathbf{Y}_t$ and $\widetilde{\mathbf{Y}}_{t-1}$.

2.3. *Estimation.* Given a series of observations $\mathbf{Y}_{-m+1}, \ldots, \mathbf{Y}_0, \mathbf{Y}_1, \ldots, \mathbf{Y}_T$ we form sample covariance matrices

$$\mathbf{S}_{YY} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{Y}_t\mathbf{Y}_t', \qquad \widetilde{\mathbf{S}}_{--} = \frac{1}{T}\sum_{t=1}^{T}\widetilde{\mathbf{Y}}_{t-1}\widetilde{\mathbf{Y}}_{t-1}', \tag{2.9}$$

$$\mathbf{S}_{Y-} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{Y}_t\widetilde{\mathbf{Y}}_{t-1}' = \mathbf{S}_{-Y}', \qquad \widehat{\mathbf{B}} = \mathbf{S}_{Y-}\widetilde{\mathbf{S}}_{--}^{-1}. \tag{2.10}$$

The estimator $\widehat{\mathbf{B}}$ is the least squares estimator of $\mathbf{B}$ and is maximum likelihood if the $\mathbf{Z}_t$'s are normally distributed conditional on $\mathbf{Y}_{-m+1}, \ldots, \mathbf{Y}_0$ given.

The sample canonical correlations ($r_1 > r_2 > \cdots > r_p$) and vectors are defined by

$$\begin{vmatrix} -r\mathbf{S}_{YY} & \mathbf{S}_{Y-} \\ \mathbf{S}_{-Y} & -r\widetilde{\mathbf{S}}_{--} \end{vmatrix} = 0, \tag{2.11}$$

$$\begin{bmatrix} -r\mathbf{S}_{YY} & \mathbf{S}_{Y-} \\ \mathbf{S}_{-Y} & -r\widetilde{\mathbf{S}}_{--} \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{w} \end{bmatrix} = \mathbf{0}, \qquad \mathbf{a}'\mathbf{S}_{YY}\mathbf{a} = 1, \mathbf{w}'\widetilde{\mathbf{S}}_{--}\mathbf{w} = 1, \tag{2.12}$$

and $a_{ii} > 0$. These are the maximum likelihood estimators if the $\mathbf{Z}_t$'s are normally distributed. From (2.12) we obtain $\mathbf{S}_{Y-}\widetilde{\mathbf{S}}_{--}^{-1}\mathbf{S}_{-Y}\mathbf{a} = r^2\mathbf{S}_{YY}\mathbf{a}$, which is the sample analog of (2.7). This equation can be rearranged to give

$$\mathbf{S}_{Y-}\widetilde{\mathbf{S}}_{--}^{-1}\mathbf{S}_{-Y}\mathbf{a} = t(\mathbf{S}_{YY} - \mathbf{S}_{Y-}\widetilde{\mathbf{S}}_{--}^{-1}\mathbf{S}_{-Y})\mathbf{a}, \tag{2.13}$$

where $t = r^2/(1-r^2)$. Note that $\mathbf{S}_{YY} - \mathbf{S}_{Y-}\widetilde{\mathbf{S}}_{--}^{-1}\mathbf{S}_{-Y} = \mathbf{S}_{ZZ} - \mathbf{S}_{Z-}\widetilde{\mathbf{S}}_{--}^{-1}\mathbf{S}_{-Z}$ and $\sqrt{T}\mathbf{S}_{Z-}\widetilde{\mathbf{S}}^{-1}\mathbf{S}_{-Z} \xrightarrow{p} \mathbf{0}$ as $T \to \infty$. Hence, $\mathbf{S}_{YY} - \mathbf{S}_{Y-}\widetilde{\mathbf{S}}_{--}^{-1}\mathbf{S}_{-Y}$ is asymptotically equivalent to $\mathbf{S}_{ZZ}$. In what follows we shall not distinguish between $\mathbf{S}_{ZZ} = T^{-1}\sum_{t=1}^{T}\mathbf{Z}_t\mathbf{Z}_t'$ and $\mathbf{S}_{YY} - \mathbf{S}_{Y-}\widetilde{\mathbf{S}}_{--}^{-1}\mathbf{S}_{-Y} = T^{-1}\sum_{t=1}^{T}\widehat{\mathbf{Z}}_t\widehat{\mathbf{Z}}_t'$, where $\widehat{\mathbf{Z}}_t = \mathbf{Y}_t - \widehat{\mathbf{B}}\widetilde{\mathbf{Y}}_{t-1}$ is the residual.

When $m = 1$, $\widetilde{\mathbf{Y}}_t = \mathbf{Y}_t$, $\widetilde{\mathbf{\Gamma}}_h = \mathbf{\Gamma}_h$, etc. Let $\mathbf{\Gamma}_0 = \mathbf{\Gamma}$. We define

$$\mathbf{S}_{--} = \widetilde{\mathbf{S}}_{--} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{Y}_{t-1}\mathbf{Y}_{t-1}', \qquad \mathbf{S}_{-Z} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{Y}_{t-1}\mathbf{Z}_t'. \tag{2.14}$$

Section 6 studies these problems for $m > 1$.

3. Asymptotic distribution of sample matrices. To find the asymptotic distribution of the canonical correlations $r_1, \ldots, r_p$ and canonical variates $\mathbf{a}_1, \ldots, \mathbf{a}_p$ and $\mathbf{w}_1, \ldots, \mathbf{w}_p$ we need the asymptotic distribution of $\mathbf{S}_{YY}$, $\mathbf{S}_{Y-}$ and $\widetilde{\mathbf{S}}_{--}$ or of $\mathbf{S}_{Z-}$, $\widetilde{\mathbf{S}}_{--}$ and $\mathbf{S}_{ZZ}$.

If $\mathbf{Z}_t$ has finite fourth moments, then $\sqrt{T}(\mathbf{S}_{YY} - \mathbf{\Gamma}_0)$ has a limiting normal distribution [Anderson (1971)]. Further, if $\mathbf{Z}_t$ is normally distributed, the fourth-order moments of $\mathbf{Y}_t$ and the second-order moments of $\mathbf{S}_{YY}$ are quadratic functions

of $\mathbf{\Gamma}$. In this section we find the asymptotic covariances of sample matrices when $m = 1$. The asymptotic covariances when $m > 1$ can be obtained from the application of $\widetilde{\mathbf{Y}}_t$.

To express the covariances of the sample matrices we use the "vec" notation. For $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_n)$ we define $\operatorname{vec}\mathbf{A} = (\mathbf{a}_1', \ldots, \mathbf{a}_n')'$. The Kronecker product of two matrices $\mathbf{A} = (\mathbf{a}_{ij})$ and $\mathbf{B}$ is $\mathbf{A} \otimes \mathbf{B} = (\mathbf{a}_{ij}\mathbf{B})$. A basic relation is $\operatorname{vec}\mathbf{ABC} = (\mathbf{C}' \otimes \mathbf{A})\operatorname{vec}\mathbf{B}$, which implies $\operatorname{vec}\mathbf{x}\mathbf{y}' = \operatorname{vec}\mathbf{x}1\mathbf{y}' = (\mathbf{y} \otimes \mathbf{x})\operatorname{vec} 1 = \mathbf{y} \otimes \mathbf{x}$. Define the commutator matrix $\mathbf{K}$ as the (square) permutation matrix such that $\operatorname{vec}\mathbf{A}' = \mathbf{K}\operatorname{vec}\mathbf{A}$ for every square matrix of the same order as $\mathbf{K}$. Note that $\mathbf{K}(\mathbf{A} \otimes \mathbf{B}) = (\mathbf{B} \otimes \mathbf{A})\mathbf{K}$.

THEOREM 1. *If the $\mathbf{Z}_t$'s are independently normally distributed, the limiting distribution of* $\operatorname{vec}\mathbf{S}_{ZZ}^* = \sqrt{T}\operatorname{vec}(\mathbf{S}_{ZZ} - \mathbf{\Sigma})$, $\operatorname{vec}\mathbf{S}_{-Z}^* = \sqrt{T}\operatorname{vec}\mathbf{S}_{-Z}$, *and* $\operatorname{vec}\mathbf{S}_{--}^* = \sqrt{T}\operatorname{vec}(\mathbf{S}_{--} - \mathbf{\Gamma})$ *is normal with means* $\mathbf{0}$, $\mathbf{0}$ *and* $\mathbf{0}$ *and covariances*

$$\mathcal{E}\operatorname{vec}\mathbf{S}_{ZZ}^*(\operatorname{vec}\mathbf{S}_{ZZ}^*)' = (\mathbf{I} + \mathbf{K})(\mathbf{\Sigma} \otimes \mathbf{\Sigma}), \tag{3.1}$$

$$\mathcal{E}\operatorname{vec}\mathbf{S}_{-Z}^*(\operatorname{vec}\mathbf{S}_{-Z}^*)' = \mathbf{\Sigma} \otimes \mathbf{\Gamma}, \tag{3.2}$$

$$\mathcal{E}\operatorname{vec}\mathbf{S}_{-Z}^*(\operatorname{vec}\mathbf{S}_{ZZ}^*)' = \mathbf{0}, \tag{3.3}$$

$$\mathcal{E}\operatorname{vec}\mathbf{S}_{--}^*(\operatorname{vec}\mathbf{S}_{ZZ}^*)' \to (\mathbf{I} + \mathbf{K})[\mathbf{I} - (\mathbf{B} \otimes \mathbf{B})]^{-1}(\mathbf{\Sigma} \otimes \mathbf{\Sigma}), \tag{3.4}$$

$$\mathcal{E}\operatorname{vec}\mathbf{S}_{--}^*(\operatorname{vec}\mathbf{S}_{-Z}^*)' \to (\mathbf{I} + \mathbf{K})[\mathbf{I} - (\mathbf{B} \otimes \mathbf{B})]^{-1}(\mathbf{\Sigma} \otimes \mathbf{B}\mathbf{\Gamma}), \tag{3.5}$$

$$\begin{aligned}
&\mathcal{E}\operatorname{vec}\mathbf{S}_{--}^*(\operatorname{vec}\mathbf{S}_{--}^*)' \\
&\quad\to (\mathbf{I} + \mathbf{K})[\mathbf{I} - (\mathbf{B} \otimes \mathbf{B})]^{-1}[(\mathbf{\Gamma} \otimes \mathbf{\Sigma}) + (\mathbf{\Sigma} \otimes \mathbf{\Gamma}) - (\mathbf{\Sigma} \otimes \mathbf{\Sigma})] \\
&\qquad\times [\mathbf{I} - (\mathbf{B}' \otimes \mathbf{B}')]^{-1} \\
&\quad= (\mathbf{I} + \mathbf{K})\big\{[\mathbf{I} - (\mathbf{B} \otimes \mathbf{B})]^{-1}(\mathbf{\Gamma} \otimes \mathbf{\Gamma}) \\
&\qquad\qquad + (\mathbf{\Gamma} \otimes \mathbf{\Gamma})[\mathbf{I} - (\mathbf{B}' \otimes \mathbf{B}')]^{-1} - (\mathbf{\Gamma} \otimes \mathbf{\Gamma})\big\}.
\end{aligned} \tag{3.6}$$

First we prove the following lemma.

LEMMA 1. *If $\{\mathbf{Z}_t\}$ consists of uncorrelated random vectors with $\mathcal{E}\mathbf{Z}_t = \mathbf{0}$ and $\mathcal{E}\mathbf{Z}_t\mathbf{Z}_t' = \mathbf{\Sigma}$, then*

$$\mathbf{S}_{--} - \mathbf{B}\mathbf{S}_{--}\mathbf{B}' = \mathbf{B}\mathbf{S}_{-Z} + \mathbf{S}_{Z-}\mathbf{B}' + \mathbf{S}_{ZZ} + O_p\left(\frac{1}{T}\right) \tag{3.7}$$

as $T \to \infty$.

PROOF. The model $\mathbf{Y}_t = \mathbf{B}\mathbf{Y}_{t-1} + \mathbf{Z}_t$ implies

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{Y}_t\mathbf{Y}_t' = \mathbf{B}\mathbf{S}_{--}\mathbf{B}' + \mathbf{B}\mathbf{S}_{-Z} + \mathbf{S}_{Z-}\mathbf{B}' + \mathbf{S}_{ZZ}. \tag{3.8}$$

Since $\sum_{t=1}^{T} \mathbf{Y}_t\mathbf{Y}_t' = T\mathbf{S}_{--} + \mathbf{Y}_T\mathbf{Y}_T' - \mathbf{Y}_0\mathbf{Y}_0'$, Lemma 1 follows. □

PROOF OF THEOREM 1. The sample covariances of a stationary autoregressive process are asymptotically normally distributed; see Anderson (1971), Chapter 5, for example. It remains to prove (3.1) to 3.6). If $\mathbf{X} \sim N(\mathbf{0}, \mathbf{\Sigma})$, $\mathcal{E}X_iX_jX_kX_l = \sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}$; in vec notation $\mathcal{E}\operatorname{vec}\mathbf{X}\mathbf{X}'(\operatorname{vec}\mathbf{X}\mathbf{X}')' = \mathcal{E}(\mathbf{X}\otimes\mathbf{X})(\mathbf{X}'\otimes\mathbf{X}') = \operatorname{vec}\mathbf{\Sigma}(\operatorname{vec}\mathbf{\Sigma})' + (\mathbf{\Sigma}\otimes\mathbf{\Sigma}) + \mathbf{K}(\mathbf{\Sigma}\otimes\mathbf{\Sigma})$. If $\mathbf{X}$ and $\mathbf{Y}$ are independent, $\mathcal{E}\operatorname{vec}\mathbf{X}\mathbf{X}'(\operatorname{vec}\mathbf{Y}\mathbf{Y}')' = \mathcal{E}\mathbf{X}\mathbf{X}'\otimes\mathcal{E}\mathbf{Y}\mathbf{Y}'$,

$$\begin{aligned}\mathcal{E}\operatorname{vec}\mathbf{X}\mathbf{Y}'(\operatorname{vec}\mathbf{X}\mathbf{Y}')' &= \mathcal{E}(\mathbf{Y}\otimes\mathbf{X})(\mathbf{Y}'\otimes\mathbf{X}')\\ &= \mathcal{E}(\mathbf{Y}\mathbf{Y}'\otimes\mathbf{X}\mathbf{X}') = \mathcal{E}\mathbf{Y}\mathbf{Y}'\otimes\mathcal{E}\mathbf{X}\mathbf{X}',\end{aligned} \tag{3.9}$$

$$\mathcal{E}\operatorname{vec}\mathbf{Y}\mathbf{X}'(\operatorname{vec}\mathbf{X}\mathbf{Y}')' = \mathbf{K}\mathcal{E}\operatorname{vec}\mathbf{X}\mathbf{Y}'(\operatorname{vec}\mathbf{X}\mathbf{Y}')' = \mathbf{K}(\mathcal{E}\mathbf{Y}\mathbf{Y}'\otimes\mathcal{E}\mathbf{X}\mathbf{X}'). \tag{3.10}$$

The asymptotic variance of $\operatorname{vec}\mathbf{S}_{ZZ}^* = \operatorname{vec}(1/\sqrt{T})\sum_{t=1}^{T}(\mathbf{Z}_t\mathbf{Z}_t' - \mathbf{\Sigma})$ is (3.1). To show (3.2) we write

$$\begin{aligned}\mathcal{E}\operatorname{vec}\mathbf{S}_{-Z}^*(\operatorname{vec}\mathbf{S}_{-Z}^*)' &= \frac{1}{T}\mathcal{E}\sum_{t,s=1}^{T}(\mathbf{Z}_t\otimes\mathbf{Y}_{t-1})(\mathbf{Z}_s'\otimes\mathbf{Y}_{s-1}')\\ &= \frac{1}{T}\mathcal{E}\sum_{t,s=1}^{T}(\mathbf{Z}_t\mathbf{Z}_s'\otimes\mathbf{Y}_{t-1}\mathbf{Y}_{s-1}').\end{aligned} \tag{3.11}$$

Next, (3.3) follows from $\sqrt{T}\operatorname{vec}\mathbf{S}_{-Z}^* = \sum_{t=1}^{T}(\mathbf{Z}_t\otimes\mathbf{Y}_{t-1})$ and $\sqrt{T}\operatorname{vec}\mathbf{S}_{ZZ}^* = (1/\sqrt{T})\sum_{s=1}^{T}[(\mathbf{Z}_s\otimes\mathbf{Z}_s) - \operatorname{vec}\mathbf{\Sigma}]$. Then (3.4), (3.5) and (3.6) are consequences of (3.7), (3.1), (3.2) and (3.3) and

$$\begin{aligned}\operatorname{vec}\mathbf{S}_{--}^* &= [\mathbf{I} - (\mathbf{B}\otimes\mathbf{B})]^{-1}\{(\mathbf{I}\otimes\mathbf{B})\operatorname{vec}\mathbf{S}_{-Z}^* + (\mathbf{B}\otimes\mathbf{I})\operatorname{vec}\mathbf{S}_{Z-}^* + \operatorname{vec}\mathbf{S}_{ZZ}^*\}\\ &\quad + O\left(\frac{1}{T}\right)\\ &= [\mathbf{I} - (\mathbf{B}\otimes\mathbf{B})]^{-1}\{(\mathbf{I}+\mathbf{K})(\mathbf{I}\otimes\mathbf{B})\operatorname{vec}\mathbf{S}_{-Z}^* + \operatorname{vec}\mathbf{S}_{ZZ}^*\}\\ &\quad + O\left(\frac{1}{T}\right).\end{aligned} \tag{3.12}$$

The second form of (3.6) follows from the substitution of $\mathbf{\Sigma} = \mathbf{\Gamma} - \mathbf{B}\mathbf{\Gamma}\mathbf{B}'$ in the first form. □

If $\mathbf{B} = \mathbf{0}$, then $\mathbf{Y}_t = \mathbf{Z}_t$, $\mathbf{\Gamma} = \mathbf{\Sigma}$ and (3.6) reduces to (3.1). The effect of $\mathbf{B} \neq \mathbf{0}$ is to tend to inflate the asymptotic covariance matrix of $\operatorname{vec}\mathbf{S}_{--}$. The characteristic roots of $\mathbf{B}$ and hence of $\mathbf{B}\otimes\mathbf{B}$ (products of the roots of $\mathbf{B}$) are less than 1 in absolute value; hence $|\mathbf{I} - (\mathbf{B}\otimes\mathbf{B})| \neq 0$. If $p = 1$, $\mathbf{B}\otimes\mathbf{B} = b_{11}^2$ and $\mathbf{I} - (\mathbf{B}\otimes\mathbf{B}) = 1 - b_{11}^2$.

In the classical regression model $\mathbf{Y}_t = \mathbf{B}\mathbf{X}_t + \mathbf{Z}_t$ with $\mathcal{E}\mathbf{X}_t = \mathbf{0}$, $\mathcal{E}\mathbf{Z}_t = \mathbf{0}$, $\mathcal{E}\mathbf{X}_t\mathbf{X}_t' = \mathbf{\Sigma}_{XX}$, $\mathcal{E}\mathbf{Z}_t\mathbf{Z}_t' = \mathbf{\Sigma}$, and $\mathcal{E}\mathbf{X}_t\mathbf{Z}_t' = \mathbf{0}$. Then $\mathcal{E}\mathbf{Y}_t = \mathbf{0}$ and $\mathcal{E}\mathbf{Y}_t\mathbf{Y}_t' = \mathbf{B}\mathbf{\Sigma}_{XX}\mathbf{B}' + \mathbf{\Sigma} = \mathbf{\Sigma}_{YY}$. If the $\mathbf{X}_t$'s and $\mathbf{Z}_t$'s are independent,

$$\operatorname{Cov}[\operatorname{vec}\mathbf{S}_{YY}, \operatorname{vec}\mathbf{S}_{YY}] = (\mathbf{I}+\mathbf{K})(\mathbf{\Sigma}_{YY}\otimes\mathbf{\Sigma}_{YY}). \tag{3.13}$$

In the autoregressive model $\mathcal{E}\mathbf{Y}_t\mathbf{Y}_t' = \mathbf{\Gamma} = \mathbf{\Sigma}_{YY}$, but (3.6) is very different from (3.13). In the scalar case (3.6) is $2[(1+\beta^2)/(1-\beta^2)]\sigma^2$ instead of $2\sigma^2$, where $\sigma^2 = \mathcal{E}y_t^2$; the factor $(1+\beta^2)/(1-\beta^2)$ inflates the variance of $\sum_{t=1}^T y_t^2/\sqrt{T}$. The larger β^2 is, the larger the inflation factor.

Note that the covariances depend on $\mathbf{Z}_t$ being normal.

4. Asymptotic distributions of canonical correlations and vectors. Let the solutions to

$$\mathbf{B\Gamma B}'\boldsymbol{\phi} = \theta\mathbf{\Sigma}\boldsymbol{\phi}, \qquad \boldsymbol{\phi}'\mathbf{\Sigma}\boldsymbol{\phi} = 1 \tag{4.1}$$

form the matrices

$$\mathbf{\Phi} = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_p), \qquad \mathbf{\Theta} = \operatorname{diag}(\theta_1, \ldots, \theta_p) \tag{4.2}$$

with $\theta_1 \geq \theta_2 \geq \cdots \geq \theta_p \geq 0$ and $\phi_{ii} > 0$, $i = 1, \ldots, p$. Let the solutions to

$$\widehat{\mathbf{B}}\mathbf{S}_{--}\widehat{\mathbf{B}}'\mathbf{f} = t\mathbf{S}_{ZZ}\mathbf{f}, \qquad \mathbf{f}'\mathbf{S}_{ZZ}\mathbf{f} = 1 \tag{4.3}$$

form the matrices

$$\mathbf{F} = (\mathbf{f}_1, \ldots, \mathbf{f}_p), \qquad \mathbf{T} = \operatorname{diag}(t_1, \ldots, t_p) \tag{4.4}$$

with $t_1 > t_2 > \cdots > t_p$ and $f_{ii} > 0$, $i = 1, \ldots, p$. We shall find the asymptotic distribution of $\mathbf{F}$ and $\mathbf{T}$ when the $\mathbf{Z}_t$'s are normally distributed.

Since $\mathbf{\Gamma} = \mathbf{B\Gamma B}' + \mathbf{\Sigma}$, (4.1) is equivalent to

$$\mathbf{\Gamma}\boldsymbol{\phi} = \delta\mathbf{\Sigma}\boldsymbol{\phi}, \qquad \boldsymbol{\phi}'\mathbf{\Sigma}\boldsymbol{\phi} = 1, \tag{4.5}$$

where $\delta = (1-\rho^2)^{-1} = \theta + 1$. Similarly, (4.3) is asymptotically equivalent to

$$\mathbf{S}_{--}\mathbf{f} = d\mathbf{S}_{ZZ}\mathbf{f}, \qquad \mathbf{f}'\mathbf{S}_{ZZ}\mathbf{f} = 1, \tag{4.6}$$

where $d = (1-r^2)^{-1} = t + 1$, since

$$\begin{aligned}\sqrt{T}(\mathbf{S}_{YY} - \mathbf{S}_{--}) &= \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^T \mathbf{Y}_t\mathbf{Y}_t' - \frac{1}{T}\sum_{t=1}^T \mathbf{Y}_{t-1}\mathbf{Y}_{t-1}'\right) \\ &= \frac{1}{\sqrt{T}}(\mathbf{Y}_T\mathbf{Y}_T' - \mathbf{Y}_0\mathbf{Y}_0') \xrightarrow{p} \mathbf{0}.\end{aligned} \tag{4.7}$$

Let $\mathbf{\Phi}'\mathbf{Y}_t = \mathbf{X}_t$, $\mathbf{\Phi}'\mathbf{Z}_t = \mathbf{W}_t$ and $\mathbf{\Phi}'\mathbf{B}(\mathbf{\Phi}')^{-1} = \mathbf{\Psi}$. Then $\mathbf{X}_t$ satisfies

$$\mathbf{X}_t = \mathbf{\Psi}\mathbf{X}_{t-1} + \mathbf{W}_t, \tag{4.8}$$

$$\mathcal{E}\mathbf{X}_t\mathbf{X}_t' = \mathbf{\Phi}'\mathbf{\Gamma\Phi} = \mathbf{\Theta} + \mathbf{I} = \mathbf{\Delta} = \operatorname{diag}(\delta_1, \ldots, \delta_p), \tag{4.9}$$

$$\mathcal{E}\mathbf{W}_t\mathbf{W}_t' = \mathbf{\Phi}'\mathbf{\Sigma\Phi} = \mathbf{I} \tag{4.10}$$

and $\mathcal{E}\mathbf{X}_{t-1}\mathbf{W}_t' = \mathbf{0}$. Let

$$\mathbf{T}_{--} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_{t-1}\mathbf{X}_{t-1}', \tag{4.11}$$

$$\mathbf{T}_{WW} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{W}_t\mathbf{W}_t'. \tag{4.12}$$

The asymptotic covariances of $\mathbf{T}_{--}$ and $\mathbf{T}_{WW}$ are given in Theorem 1 with $\mathbf{S}_{--}$ and $\mathbf{S}_{ZZ}$ replaced by $\mathbf{T}_{--}$ and $\mathbf{T}_{WW}$, $\mathbf{\Sigma}$ by $\mathbf{I}$, $\mathbf{\Gamma}$ by $\mathbf{\Delta}$ and $\mathbf{B}$ by $\mathbf{\Psi}$.

THEOREM 2. *If the $\mathbf{Z}_t$'s are independently normally distributed, the limiting distribution of $\mathbf{T}_{--}^* = \sqrt{T}(\mathbf{T}_{--} - \mathbf{\Delta})$ and $\mathbf{T}_{WW}^* = \sqrt{T}(\mathbf{T}_{WW} - \mathbf{I})$ is normal with means $\mathbf{0}$ and $\mathbf{0}$ and covariances*

$$\begin{aligned}\mathcal{E}\,\mathrm{vec}\,\mathbf{T}_{--}^*(\mathrm{vec}\,\mathbf{T}_{--}^*)' &\to (\mathbf{I}+\mathbf{K})[\mathbf{I}-(\mathbf{\Psi}\otimes\mathbf{\Psi})]^{-1}\\ &\quad\times[(\mathbf{\Delta}\otimes\mathbf{I})+(\mathbf{I}\otimes\mathbf{\Delta})-(\mathbf{I}\otimes\mathbf{I})][\mathbf{I}-(\mathbf{\Psi}'\times\mathbf{\Psi}')]^{-1}\\ &= (\mathbf{I}+\mathbf{K})\big\{[\mathbf{I}-(\mathbf{\Psi}\otimes\mathbf{\Psi})]^{-1}(\mathbf{\Delta}\otimes\mathbf{\Delta})\\ &\qquad+(\mathbf{\Delta}\otimes\mathbf{\Delta})[\mathbf{I}-(\mathbf{\Psi}'\otimes\mathbf{\Psi}')]^{-1}-(\mathbf{\Delta}\otimes\mathbf{\Delta})\big\},\end{aligned} \tag{4.13}$$

$$\mathcal{E}\,\mathrm{vec}\,\mathbf{T}_{WW}^*(\mathrm{vec}\,\mathbf{T}_{WW}^*)' = (\mathbf{I}+\mathbf{K})(\mathbf{I}\otimes\mathbf{I}), \tag{4.14}$$

$$\mathcal{E}\,\mathrm{vec}\,\mathbf{T}_{--}^*(\mathrm{vec}\,\mathbf{T}_{WW}^*)' \to (\mathbf{I}+\mathbf{K})[\mathbf{I}-(\mathbf{\Psi}\otimes\mathbf{\Psi})]^{-1}(\mathbf{I}\otimes\mathbf{I}). \tag{4.15}$$

Note that $(\mathbf{\Delta}\otimes\mathbf{I})$, $(\mathbf{I}\otimes\mathbf{\Delta})$ and $(\mathbf{I}\otimes\mathbf{I})$ are diagonal.

Let $\mathbf{h} = \mathbf{\Phi}^{-1}\mathbf{f}$, $\mathbf{H} = \mathbf{\Phi}^{-1}\mathbf{F}$, $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_p)$. Then $\mathbf{H}$ and $\mathbf{D}$ satisfy

$$\mathbf{T}_{--}\mathbf{H} = \mathbf{T}_{WW}\mathbf{H}\mathbf{D}, \qquad \mathbf{H}'\mathbf{T}_{WW}\mathbf{H} = \mathbf{I}. \tag{4.16}$$

Since $\mathbf{T}_{--} \xrightarrow{p} \mathbf{\Delta}$ and $\mathbf{T}_{WW} \xrightarrow{p} \mathbf{I}$, the probability limit of (4.16) is $\mathbf{\Delta}\mathbf{H}_\infty = \mathbf{H}_\infty\mathbf{\Delta}$, $\mathbf{H}_\infty'\mathbf{H}_\infty = \mathbf{I}$, which implies $\mathrm{plim}_{T\to\infty}\mathbf{H}$ is diagonal with $\mathrm{plim}_{T\to\infty} h_{ii} = \pm 1$ if the diagonal elements of $\mathbf{\Delta}$ are different. Since $\phi_{ii} > 0$ and $\mathrm{plim}_{T\to\infty} f_{ii} > 0$, then $\mathrm{plim}_{T\to\infty} h_{ii} > 0$. Hence $\mathbf{H} \xrightarrow{p} \mathbf{I}$ and $\mathbf{D} \xrightarrow{p} \mathbf{\Delta}$.

Let $\mathbf{H}^* = \sqrt{T}(\mathbf{H} - \mathbf{I})$, and $\mathbf{D}^* = \sqrt{T}(\mathbf{D} - \mathbf{\Delta}) = \sqrt{T}(\mathbf{T} - \mathbf{\Theta})$, where $\mathbf{T} = \mathrm{diag}(t_1, \ldots, t_p)$. From (4.16) we obtain

$$\mathbf{H}^*\mathbf{\Delta} - \mathbf{\Delta}\mathbf{H}^* + \mathbf{D}^* = \mathbf{T}_{--}^* - \mathbf{T}_{WW}^*\mathbf{\Delta} + o_p(1), \tag{4.17}$$

$$\mathbf{H}^{*\prime} + \mathbf{H}^* = -\mathbf{T}_{WW}^* + o_p(1). \tag{4.18}$$

In components (4.17) and (4.18) are $h_{ij}^*(\delta_j - \delta_i) = t_{ij}^{*--} - t_{ij}^{*WW}\delta_j + o_p(1)$, $i \neq j$, $d_{ii}^* = t_{ii}^{*--} - t_{ii}^{*WW}\delta_i + o_p(1)$, $2h_{ii}^* = -t_{ii}^{*WW} + o_p(1)$.

Let

$$\mathbf{E} = \sum_{i=1}^{p} (\boldsymbol{\varepsilon}_i \otimes \boldsymbol{\varepsilon}_i)(\boldsymbol{\varepsilon}_i' \otimes \boldsymbol{\varepsilon}_i'), \tag{4.19}$$

where $\boldsymbol{\varepsilon}_i$ has 1 in the ith position and 0's elsewhere; $\mathbf{E}$ is a diagonal $p^2 \times p^2$ matrix with 1 in the $[(i-1)p+i, (i-1)p+i]$th position, $i = 1, \ldots, p$, and 0's elsewhere. Then $\operatorname{vec} \mathbf{D}^* = \mathbf{E} \operatorname{vec}(\mathbf{T}_{--}^* - \mathbf{T}_{WW}^* \boldsymbol{\Delta})$. Note $\mathbf{E}^2 = \mathbf{E}$; hence the Moore–Penrose generalized inverse of $\mathbf{E}$ is $\mathbf{E}^+ = \mathbf{E}$. Let $\mathbf{H}^* = \mathbf{H}_d^* + \mathbf{H}_n^*$, where $\mathbf{H}_d^* = \operatorname{diag}(h_{11}^*, \ldots, h_{pp}^*)$. Then $\operatorname{vec} \mathbf{H}_d^* = \mathbf{E} \operatorname{vec} \mathbf{H}^*$.

To write (4.17) more suitably we note that $\operatorname{vec} \mathbf{H}^* \boldsymbol{\Delta} = \operatorname{vec} \mathbf{I} \mathbf{H}^* \boldsymbol{\Delta} = (\boldsymbol{\Delta} \otimes \mathbf{I}) \operatorname{vec} \mathbf{H}^*$ and $\operatorname{vec} \boldsymbol{\Delta} \mathbf{H}^* = \operatorname{vec} \boldsymbol{\Delta} \mathbf{H}^* \mathbf{I} = (\mathbf{I} \otimes \boldsymbol{\Delta}) \operatorname{vec} \mathbf{H}^*$. Define

$$\begin{aligned} \mathbf{N} &= (\boldsymbol{\Delta} \otimes \mathbf{I}) - (\mathbf{I} \otimes \boldsymbol{\Delta}) \\ &= \operatorname{diag}(\delta_1 \mathbf{I} - \boldsymbol{\Delta}, \delta_2 \mathbf{I} - \boldsymbol{\Delta}, \ldots, \delta_p \mathbf{I} - \boldsymbol{\Delta}) \\ &= \operatorname{diag}(0, \delta_1 - \delta_2, \ldots, \delta_1 - \delta_p, \delta_2 - \delta_1, 0, \ldots, \delta_p - \delta_{p-1}, 0). \end{aligned} \tag{4.20}$$

The Moore–Penrose inverse of $\mathbf{N}$ is

$$\begin{aligned} \mathbf{N}^+ = \operatorname{diag}\Big(0, \frac{1}{\delta_1 - \delta_2}, \ldots, \frac{1}{\delta_1 - \delta_p}, \\ \frac{1}{\delta_2 - \delta_1}, 0, \frac{1}{\delta_2 - \delta_3}, \ldots, \frac{1}{\delta_p - \delta_{p-1}}, 0\Big). \end{aligned} \tag{4.21}$$

Then $\mathbf{N}\mathbf{N}^+ = \mathbf{I}_{p^2} - \mathbf{E}$ and $\mathbf{N}\mathbf{E} = \mathbf{N}^+\mathbf{E} = \mathbf{0}$.

We can write (4.17) as

$$\begin{aligned} \mathbf{N} \operatorname{vec} \mathbf{H}^* + \operatorname{vec} \mathbf{D}^* &= \operatorname{vec}(\mathbf{T}_{--}^* - \mathbf{T}_{WW}^* \boldsymbol{\Delta}) + o_p(1) \\ &= \operatorname{vec} \mathbf{T}_{--}^* - (\boldsymbol{\Delta} \otimes \mathbf{I}) \operatorname{vec} \mathbf{T}_{WW}^* + o_p(1). \end{aligned} \tag{4.22}$$

From (4.22) we obtain

$$\operatorname{vec} \mathbf{H}_n^* = \mathbf{N}^+ \big[\operatorname{vec} \mathbf{T}_{--}^* - (\boldsymbol{\Delta} \otimes \mathbf{I}) \operatorname{vec} \mathbf{T}_{WW}^*\big] + o_p(1), \tag{4.23}$$

$$\operatorname{vec} \mathbf{D}^* = \mathbf{E} \big[\operatorname{vec} \mathbf{T}_{--}^* - (\boldsymbol{\Delta} \otimes \mathbf{I}) \operatorname{vec} \mathbf{T}_{WW}^*\big] + o_p(1). \tag{4.24}$$

In vec notation, (4.18) is $(\mathbf{I} + \mathbf{K}) \operatorname{vec} \mathbf{H}^* = -\operatorname{vec} \mathbf{T}_{WW}^* + o_p(1)$, from which we obtain

$$\mathbf{E}(\mathbf{I} + \mathbf{K}) \operatorname{vec} \mathbf{H}^* = 2\mathbf{E} \operatorname{vec} \mathbf{H}^* = 2 \operatorname{vec} \mathbf{H}_d^* = -\mathbf{E} \operatorname{vec} \mathbf{T}_{WW}^* + o_p(1). \tag{4.25}$$

DEFINITION. If $(\mathbf{X}_T, \mathbf{Y}_T) \xrightarrow{d} (\mathbf{X}, \mathbf{Y})$ with $\mathcal{E}\mathbf{X}'\mathbf{X} < \infty$ and $\mathcal{E}\mathbf{Y}'\mathbf{Y} < \infty$, then

$$A \operatorname{Cov}(\mathbf{X}_T, \mathbf{Y}_T) = \mathcal{E}(\mathbf{X} - \mathcal{E}\mathbf{X})(\mathbf{Y} - \mathcal{E}\mathbf{Y})'. \tag{4.26}$$

The asymptotic covariance matrix of $\operatorname{vec}\mathbf{T}^*_{--} - (\boldsymbol{\Delta}\otimes\mathbf{I})\operatorname{vec}\mathbf{T}^*_{WW}$ is

$$
\begin{aligned}
&A\operatorname{Cov}\big[\operatorname{vec}\mathbf{T}^*_{--} - (\boldsymbol{\Delta}\otimes\mathbf{I})\operatorname{vec}\mathbf{T}^*_{WW}, \operatorname{vec}\mathbf{T}^*_{--} - (\boldsymbol{\Delta}\otimes\mathbf{I})\operatorname{vec}\mathbf{T}^*_{WW}\big]\\
&\quad= (\mathbf{I}+\mathbf{K})[\mathbf{I}-(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi})]^{-1}(\boldsymbol{\Delta}\otimes\boldsymbol{\Theta})\\
&\qquad+(\boldsymbol{\Delta}\otimes\boldsymbol{\Theta})[\mathbf{I}-(\boldsymbol{\Psi}'\otimes\boldsymbol{\Psi}')]^{-1}(\mathbf{I}+\mathbf{K})+(\boldsymbol{\Delta}^2\otimes\mathbf{I})-(\boldsymbol{\Delta}\otimes\boldsymbol{\Delta})\\
&\quad= (\mathbf{I}+\mathbf{K})\boldsymbol{\Lambda}(\boldsymbol{\Delta}\otimes\boldsymbol{\Theta})+(\boldsymbol{\Delta}\otimes\boldsymbol{\Theta})\boldsymbol{\Lambda}'(\mathbf{I}+\mathbf{K})+\mathbf{N}^+(\boldsymbol{\Delta}\otimes\mathbf{I}),
\end{aligned}
\tag{4.27}
$$

where

$$
\boldsymbol{\Lambda} = [\mathbf{I}-(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi})]^{-1}. \tag{4.28}
$$

The element of $\boldsymbol{\Lambda}$ in the jth row of the ith block of rows and in the ℓth column of the kth block of columns is denoted as $\lambda_{ij,k\ell}$. Note $\boldsymbol{\Theta} = \boldsymbol{\Delta} - \mathbf{I}$.

The asymptotic covariance matrix of $\operatorname{vec}\mathbf{T}^*_{WW}$ is (4.14), and the asymptotic covariance matrix between $\operatorname{vec}\mathbf{T}^*_{--} - (\boldsymbol{\Delta}\otimes\mathbf{I})\operatorname{vec}\mathbf{T}^*_{WW}$ and $\operatorname{vec}\mathbf{T}^*_{WW}$ is

$$
\begin{aligned}
&A\operatorname{Cov}\big[\operatorname{vec}\mathbf{T}^*_{--} - (\boldsymbol{\Delta}\otimes\mathbf{I})\operatorname{vec}\mathbf{T}^*_{WW}, \operatorname{vec}\mathbf{T}^*_{WW}\big]\\
&\quad= (\mathbf{I}+\mathbf{K})[\mathbf{I}-(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi})]^{-1}(\mathbf{I}\otimes\mathbf{I})-(\boldsymbol{\Delta}\otimes\mathbf{I})(\mathbf{I}+\mathbf{K})(\mathbf{I}\otimes\mathbf{I})\\
&\quad= (\mathbf{I}+\mathbf{K})\boldsymbol{\Lambda}-(\boldsymbol{\Delta}\otimes\mathbf{I})(\mathbf{I}+\mathbf{K}).
\end{aligned}
\tag{4.29}
$$

The asymptotic covariance matrices of $\operatorname{vec}\mathbf{H}^*_n$, $\operatorname{vec}\mathbf{H}^*_d$ and $\operatorname{vec}\mathbf{D}^*$ are found from (4.29), (4.14) and (4.27).

From (4.24), (4.27), $\mathbf{EK} = \mathbf{E}$, $\mathbf{E}(\boldsymbol{\Delta}^2\otimes\mathbf{I}) = \mathbf{E}(\boldsymbol{\Delta}\otimes\boldsymbol{\Delta})$, $\mathbf{E}(\boldsymbol{\Delta}\otimes\mathbf{I}) = \mathbf{E}(\mathbf{I}\otimes\boldsymbol{\Delta})$ and $(\mathbf{A}\otimes\mathbf{B})\mathbf{K} = \mathbf{K}(\mathbf{B}\otimes\mathbf{A})$, we obtain the following theorem.

THEOREM 3. *If the $\mathbf{Z}_t$'s are independently normally distributed and if the roots of $|\boldsymbol{\Gamma}-\delta\boldsymbol{\Sigma}| = 0$ are distinct, the nonzero elements of $\mathbf{D}^*$ have a limiting normal distribution with means $\mathbf{0}$; the covariance matrix of this limiting distribution is given by*

$$
\begin{aligned}
A\operatorname{Cov}(\operatorname{vec}\mathbf{D}^*, \operatorname{vec}\mathbf{D}^*) &= 2\mathbf{E}\big\{[\mathbf{I}-(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi})]^{-1}(\boldsymbol{\Delta}\otimes\boldsymbol{\Theta})\\
&\qquad+(\boldsymbol{\Delta}\otimes\boldsymbol{\Theta})[\mathbf{I}-(\boldsymbol{\Psi}'\otimes\boldsymbol{\Psi}')]^{-1}\big\}\mathbf{E}\\
&= 2\mathbf{E}[\boldsymbol{\Lambda}(\boldsymbol{\Delta}\otimes\boldsymbol{\Theta})+(\boldsymbol{\Delta}\otimes\boldsymbol{\Theta})\boldsymbol{\Lambda}']\mathbf{E}.
\end{aligned}
\tag{4.30}
$$

The asymptotic covariance of d_i and d_j is

$$
\begin{aligned}
A\operatorname{Cov}(d^*_i, d^*_j) &= 2(\lambda_{ii,jj}\delta_j\theta_j + \delta_i\theta_i\lambda_{jj,ii})\\
&= 2\left(\lambda_{ii,,jj}\frac{\rho_j^2}{(1-\rho_j^2)^2} + \lambda_{jj,ii}\frac{\rho_i^2}{(1-\rho_i^2)^2}\right)\\
&= 2(\boldsymbol{\phi}_i'\otimes\boldsymbol{\phi}_i')\big\{[\mathbf{I}-(\mathbf{B}\otimes\mathbf{B})]^{-1}[\boldsymbol{\Gamma}\otimes(\boldsymbol{\Gamma}-\boldsymbol{\Sigma})]\\
&\qquad+[\boldsymbol{\Gamma}\otimes(\boldsymbol{\Gamma}-\boldsymbol{\Sigma})][\mathbf{I}-(\mathbf{B}'\otimes\mathbf{B}')]^{-1}\big\}(\boldsymbol{\phi}_j\otimes\boldsymbol{\phi}_j).
\end{aligned}
\tag{4.31}
$$

The vector $\boldsymbol{\phi}_i$ is estimated consistently by $\mathbf{f}_i$; the matrices $\mathbf{B}$, $\boldsymbol{\Gamma}$, and $\boldsymbol{\Sigma}$ by $\widehat{\mathbf{B}}$, $\mathbf{S}_{--}$ and $\mathbf{S}_{ZZ}$, respectively, and δ_i by d_i. The variance (4.31) can, therefore, be estimated consistently.

Note that $d_i^* = t_i^* = \sqrt{T}(t_i - \theta_i)$ and $r_j^{2*} = \sqrt{T}(r_j^2 - \rho_j^2) = (1-\rho_j^2)^2\sqrt{T} \times (t_j - \theta_j) + o_p(1)$. Thus, $r_1^{2*}, \ldots, r_p^{2*}$ have a limiting normal distribution with means 0 and covariances

$$A\operatorname{Cov}(r_i^{2*}, r_j^{2*}) = 2\big[\lambda_{ii,jj}(1-\rho_i^2)^2\rho_j^2 + \lambda_{jj,ii}\rho_i^2(1-\rho_j^2)^2\big]. \tag{4.32}$$

Unless $\boldsymbol{\Psi}$ is a diagonal matrix, the roots $r_1, \ldots, r_p$ are asymptotically correlated; that is, dependent. In the special case that $\boldsymbol{\Psi}$ is diagonal ($\psi_{ij} = \psi_{ii}\delta_{ij} = \rho_i\delta_{ij}$, where $\delta_{ii} = 1$ and $\delta_{ij} = 0$, $i \neq j$), the roots are asymptotically independent, and the asymptotic variance of $d_i^* = t_i^*$ is $4\rho_i^2/(1-\rho_i^2)^3$ and the ith component of $\mathbf{X}_t = \boldsymbol{\Psi}\mathbf{X}_{t-1} + \mathbf{W}_t$ is $\mathbf{X}_{it} = \rho_i X_{i,t-1} + W_{it}$. Contrast this result with the canonical form of $\mathbf{Y}_t = \mathbf{B}\mathbf{X}_t + \mathbf{W}_t$, namely $U_{it} = \rho_i V_{it} + W_{it}$ with $\mathcal{E}U_{it}^2 = \mathcal{E}V_{it}^2 = (1-\rho_i^2)^{-1}$ and $\mathcal{E}W_{it}^2 = 1$. In this case the asymptotic variance of t_i^* is $4\rho_i^2/(1-\rho_i^2)^2$ [Anderson (1999a)]. The ratio of the variance in the autoregressive model to that in the regression model is $(1-\rho_i^2)^{-1}$, which is greater than 1. (Note that this result agrees with the asymptotic distribution of the serial correlation coefficient in a scalar first-order autoregressive process [Theorem 5.5.6 of Anderson (1971)]. The variance of the limiting distribution of $r_i^* = \sqrt{T}(r_i - \rho_i)$ when $\boldsymbol{\Psi}$ is diagonal is $1-\rho_i^2$ as compared to $(1-\rho_i^2)^2$ in the regression model.) In the classical case, the eigenvalues are asymptotically independent because $\mathbf{Y}_t$ and $\mathbf{X}_t$ are each transformed, and hence $\mathbf{B}$ can be transformed into a diagonal matrix, but in the autoregression case $\mathbf{Y}_t$ and $\mathbf{Y}_{t-1}$ are transformed by the same linear transformation.

Now consider

$$\begin{aligned}\operatorname{vec}\mathbf{H}^* &= \operatorname{vec}\mathbf{H}_n^* + \operatorname{vec}\mathbf{H}_d^* \\ &= \mathbf{N}^+\big[\operatorname{vec}\mathbf{T}_{--}^* - (\boldsymbol{\Delta}\otimes\mathbf{I})\operatorname{vec}\mathbf{T}_{WW}^*\big] - \tfrac{1}{2}\mathbf{E}\operatorname{vec}\mathbf{T}_{WW}^*.\end{aligned} \tag{4.33}$$

Then

$$\begin{aligned}&A\operatorname{Cov}(\operatorname{vec}\mathbf{H}^*, \operatorname{vec}\mathbf{H}^*) \\ &\quad= \mathbf{N}^+\big[(\mathbf{I}+\mathbf{K})\boldsymbol{\Lambda}(\boldsymbol{\Delta}\otimes\boldsymbol{\Theta}) + (\boldsymbol{\Delta}\otimes\boldsymbol{\Theta})\boldsymbol{\Lambda}'(\mathbf{I}+\mathbf{K}) + (\boldsymbol{\Delta}^2\otimes\mathbf{I}) - (\boldsymbol{\Delta}\otimes\boldsymbol{\Delta})\big]\mathbf{N}^+ \\ &\qquad - \tfrac{1}{2}\mathbf{N}^+[(\mathbf{I}+\mathbf{K})\boldsymbol{\Lambda} - (\boldsymbol{\Delta}\otimes\mathbf{I})(\mathbf{I}+\mathbf{K})]\mathbf{E} \\ &\qquad - \tfrac{1}{2}\mathbf{E}[\boldsymbol{\Lambda}'(\mathbf{I}+\mathbf{K}) - (\mathbf{I}+\mathbf{K})(\boldsymbol{\Delta}\otimes\mathbf{I})]\mathbf{N}^+ + \tfrac{1}{4}\mathbf{E}(\mathbf{I}+\mathbf{K})\mathbf{E} \\ &\quad= \mathbf{N}^+\big[(\mathbf{I}+\mathbf{K})\boldsymbol{\Lambda}(\boldsymbol{\Delta}\otimes\boldsymbol{\Theta}) + (\boldsymbol{\Delta}\otimes\boldsymbol{\Theta})\boldsymbol{\Lambda}'(\mathbf{I}+\mathbf{K})\big]\mathbf{N}^+ \\ &\qquad + \mathbf{N}^+(\boldsymbol{\Delta}\otimes\mathbf{I}) - \tfrac{1}{2}\mathbf{N}^+(\mathbf{I}+\mathbf{K})\boldsymbol{\Lambda}\mathbf{E} - \tfrac{1}{2}\mathbf{E}\boldsymbol{\Lambda}'(\mathbf{I}+\mathbf{K})\mathbf{N}^+ + \tfrac{1}{2}\mathbf{E}\end{aligned} \tag{4.34}$$

because $\mathbf{N}^+(\boldsymbol{\Delta}\otimes\mathbf{I})\mathbf{E} = \mathbf{0}$. The asymptotic covariance matrix of $\operatorname{vec}\mathbf{F} = (\mathbf{I}\otimes\boldsymbol{\Phi})\operatorname{vec}\mathbf{H}$ is (4.34) multiplied on the left-hand side by $(\mathbf{I}\otimes\boldsymbol{\Phi})$ and on the right-hand side by $(\mathbf{I}\otimes\boldsymbol{\Phi}')$.

In the classical regression model $\mathbf{Y}_t = \mathbf{B}\mathbf{X}_t + \mathbf{Z}_t$ the eigenvalues of $\mathbf{\Sigma}_{YX}\mathbf{\Sigma}_{XX}^{-1}\mathbf{\Sigma}_{XY}$ in the metric of $\mathbf{\Sigma}_{ZZ}$ are asymptotically independent when the population eigenvalues are different. In the AR(1) case the eigenvalues are asymptotically dependent. (See the comments in Section 7.1.)

5. Estimation of reduced rank regression. If the rank of $\mathbf{B}$ in $\mathbf{Y}_t = \mathbf{B}\mathbf{Y}_{t-1} + \mathbf{Z}_t$ is specified to be k $(\leq p)$, the maximum likelihood estimator of $\mathbf{B}$ under normality with $\mathbf{Y}_0$ given [Anderson (1951)] is

$$\widehat{\mathbf{B}}_k = \mathbf{S}_{ZZ}\mathbf{F}_1\mathbf{F}_1'\widehat{\mathbf{B}} = \mathbf{S}_{Y-}\widehat{\mathbf{\Omega}}_1\widehat{\mathbf{\Omega}}_1', \tag{5.1}$$

where $\mathbf{F}_1$ and $\widehat{\mathbf{\Omega}}_1$ are the first k columns of $\mathbf{F}$ and $\widehat{\mathbf{\Omega}} = (\mathbf{w}_1, \ldots, \mathbf{w}_p)$, respectively. The matrix $\widehat{\mathbf{B}}_k$ is the *reduced rank regression* coefficient matrix. In this section the asymptotic distribution of $\widehat{\mathbf{B}}_k$ will be derived without assuming normality.

In canonical terms the reduced rank regression estimator is

$$\widehat{\mathbf{\Psi}}_k = \mathbf{T}_{WW}\mathbf{H}_1\mathbf{H}_1'\widehat{\mathbf{\Psi}}, \tag{5.2}$$

where $\mathbf{H}_1$ consists of the first k columns of $\mathbf{H}$, and the least squares estimator is

$$\widehat{\mathbf{\Psi}} = \mathbf{T}_{X-}\mathbf{T}_{--}^{-1} = \mathbf{\Psi} + \mathbf{T}_{W-}\mathbf{T}_{--}^{-1}. \tag{5.3}$$

Then

$$\begin{aligned}\widehat{\mathbf{\Psi}}_k^* &= \sqrt{T}(\widehat{\mathbf{\Psi}}_k - \mathbf{\Psi}) \\ &= \left(\mathbf{T}_{WW}^*\mathbf{I}_{(k)}\mathbf{I}_{(k)}' + \mathbf{H}_1^*\mathbf{I}_{(k)}' + \mathbf{I}_{(k)}\mathbf{H}_1^{*\prime}\right)\mathbf{\Psi} + \mathbf{I}_{(k)}\mathbf{I}_{(k)}'\mathbf{T}_{W-}^*\mathbf{\Delta}^{-1} + o_p(1),\end{aligned} \tag{5.4}$$

where $\mathbf{I}_{(k)} = (\mathbf{I}_k, \mathbf{0})'$. The submatrix consisting of the last $p-k$ rows and columns of $\mathbf{\Psi}\mathbf{\Delta}\mathbf{\Psi}' = \mathbf{\Delta} - \mathbf{I}$ is $(\mathbf{\Psi}_{21} + \mathbf{\Psi}_{22})\mathbf{\Delta}(\mathbf{\Psi}_{21}' + \mathbf{\Psi}_{22}') = \mathbf{0}$, which implies $\mathbf{\Psi}_{21} = \mathbf{0}$, $\mathbf{\Psi}_{22} = \mathbf{0}$ and

$$\mathbf{\Psi} = \begin{bmatrix} \mathbf{\Psi}_{11} & \mathbf{\Psi}_{12} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{5.5}$$

Expansion of (5.4) in terms of $\mathbf{T}_{WW}^*$, $\mathbf{T}_{W-}^*$ and $\mathbf{H}^*$ gives

$$\begin{aligned}\widehat{\mathbf{\Psi}}_k^* &= \left\{ \begin{bmatrix} \mathbf{T}_{WW}^{*11} & \mathbf{T}_{WW}^{*12} \\ \mathbf{T}_{WW}^{*21} & \mathbf{T}_{WW}^{*22} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{H}_{11}^* & \mathbf{0} \\ \mathbf{H}_{21}^* & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{H}_{11}^{*\prime} & \mathbf{H}_{21}^{*\prime} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \right\} \\ &\quad \times \begin{bmatrix} \mathbf{\Psi}_{11} & \mathbf{\Psi}_{12} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \\ &\quad + \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{T}_{W-}^{*11} & \mathbf{T}_{W-}^{*12} \\ \mathbf{T}_{W-}^{*21} & \mathbf{T}_{W-}^{*22} \end{bmatrix} \begin{bmatrix} \mathbf{\Delta}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} + o_p(1) \\ &= \begin{bmatrix} (\mathbf{T}_{WW}^{*11} + \mathbf{H}_{11}^* + \mathbf{H}_{11}^{*\prime})\mathbf{\Psi}_{11} & (\mathbf{T}_{WW}^{*11} + \mathbf{H}_{11}^* + \mathbf{H}_{11}^{*\prime})\mathbf{\Psi}_{12} \\ \quad + \mathbf{T}_{W-}^{*11}\mathbf{\Delta}_1^{-1} & \quad + \mathbf{T}_{W-}^{*12} \\ (\mathbf{T}_{WW}^{*21} + \mathbf{H}_{21}^*)\mathbf{\Psi}_{11} & (\mathbf{T}_{WW}^{*21} + \mathbf{H}_{21}^*)\mathbf{\Psi}_{12} \end{bmatrix} + o_p(1).\end{aligned} \tag{5.6}$$

The first k columns of (4.17) and the upper left-hand side submatrix of (4.18) are

$$\begin{bmatrix} \mathbf{H}_{11}^{*}\boldsymbol{\Delta}_1 - \boldsymbol{\Delta}_1\mathbf{H}_{11}^{*} + \mathbf{D}_1 \\ \mathbf{H}_{21}^{*}\boldsymbol{\Delta}_1 - \mathbf{H}_{21}^{*} \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{--}^{*11} - \mathbf{T}_{WW}^{*11}\boldsymbol{\Delta}_1 \\ \mathbf{T}_{--}^{*21} - \mathbf{T}_{WW}^{*21}\boldsymbol{\Delta}_1 \end{bmatrix} + o_p(1), \tag{5.7}$$

$$\mathbf{H}_{11}^{*} + \mathbf{H}_{11}^{*\prime} = -\mathbf{T}_{WW}^{*11} + o_p(1). \tag{5.8}$$

Use of (5.7) and (5.8) in (5.6) yields

$$\widehat{\boldsymbol{\Psi}}_k^{*} = \begin{bmatrix} \mathbf{T}_{W-}^{*11}\boldsymbol{\Delta}_1^{-1} & \mathbf{T}_{W-}^{*12} \\ (\mathbf{T}_{--}^{*21} - \mathbf{T}_{WW}^{*21}) & (\mathbf{T}_{--}^{*21} - \mathbf{T}_{WW}^{*21}) \\ \quad \times (\boldsymbol{\Delta}_1 - \mathbf{I})^{-1}\boldsymbol{\Psi}_{11} & \quad \times (\boldsymbol{\Delta}_1 - \mathbf{I})^{-1}\boldsymbol{\Psi}_{12} \end{bmatrix} + o_p(1). \tag{5.9}$$

From $\mathbf{T}_{XX} = \boldsymbol{\Psi}\mathbf{T}_{--}\boldsymbol{\Psi}' + \boldsymbol{\Psi}\mathbf{T}_{-W} + \mathbf{T}_{W-}\boldsymbol{\Psi}' + \mathbf{T}_{WW}$, $\mathbf{T}_{XX} = \mathbf{T}_{--} + o_p(1/\sqrt{T})$ and (5.5) we obtain

$$\mathbf{T}_{--}^{*21} - \mathbf{T}_{WW}^{*21} = \mathbf{T}_{W-}^{*21}\boldsymbol{\Psi}_{11}' + \mathbf{T}_{W-}^{*22}\boldsymbol{\Psi}_{12}' + o_p\left(\frac{1}{\sqrt{T}}\right). \tag{5.10}$$

Then from (5.9) we derive

$$\begin{aligned} \widehat{\boldsymbol{\Psi}}_k^{*}\boldsymbol{\Delta} &= \begin{bmatrix} \mathbf{T}_{W-}^{*11} & \mathbf{T}_{W-}^{*12} \\ (\mathbf{T}_{W-}^{*21}\boldsymbol{\Psi}_{11}' + \mathbf{T}_{W-}^{*22}\boldsymbol{\Psi}_{12}') & (\mathbf{T}_{W-}^{*21}\boldsymbol{\Psi}_{11}' + \mathbf{T}_{W-}^{*22}\boldsymbol{\Psi}_{12}') \\ \quad \times (\boldsymbol{\Delta}_1 - \mathbf{I})^{-1}\boldsymbol{\Psi}_{11} & \quad \times (\boldsymbol{\Delta}_1 - \mathbf{I})^{-1}\boldsymbol{\Psi}_{12} \end{bmatrix} + o_p(1) \\ &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{T}_{W-}^{*} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}\mathbf{T}_{W-}^{*}\begin{bmatrix} \boldsymbol{\Psi}_{11}' \\ \boldsymbol{\Psi}_{12}' \end{bmatrix}(\boldsymbol{\Delta}_1 - \mathbf{I})^{-1}(\boldsymbol{\Psi}_{11}, \boldsymbol{\Psi}_{12})\boldsymbol{\Delta} \\ &\quad + o_p(1) \\ &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{T}_{W-}^{*} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}\mathbf{T}_{W-}^{*}\mathbf{M} + o_p(1), \end{aligned} \tag{5.11}$$

where

$$\mathbf{M} = \begin{bmatrix} \boldsymbol{\Psi}_{11}' \\ \boldsymbol{\Psi}_{12}' \end{bmatrix}\boldsymbol{\Theta}_1^{-1}(\boldsymbol{\Psi}_{11}, \boldsymbol{\Psi}_{12})\boldsymbol{\Delta} = \boldsymbol{\Psi}_{1\cdot}'\boldsymbol{\Theta}_1^{-1}\boldsymbol{\Psi}_{1\cdot}\boldsymbol{\Delta}, \tag{5.12}$$

$\boldsymbol{\Psi}_{1\cdot} = (\boldsymbol{\Psi}_{11}, \boldsymbol{\Psi}_{12})$ and $\boldsymbol{\Theta}_1 = \boldsymbol{\Delta}_1 - \mathbf{I}$. Note that $\mathbf{M}^2 = \mathbf{M}$ by virtue of the upper left-hand corner of $\boldsymbol{\Psi}\boldsymbol{\Delta}\boldsymbol{\Psi}' = \boldsymbol{\Theta}$. Also $\mathbf{M}'\boldsymbol{\Delta}\mathbf{M} = \mathbf{M}'\boldsymbol{\Delta} = \boldsymbol{\Delta}\mathbf{M}$. Then

$$\operatorname{vec}\left[(\widehat{\boldsymbol{\Psi}}^{*} - \widehat{\boldsymbol{\Psi}}_k^{*})\boldsymbol{\Delta}\right] = \left[(\mathbf{I} - \mathbf{M}') \otimes \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}\right]\operatorname{vec}\mathbf{T}_{W-}^{*} + o_p(1). \tag{5.13}$$

Notice that the development to this point does not involve the second-order moments of $\mathbf{T}_{--}$, $\mathbf{T}_{-W}$ and $\mathbf{T}_{WW}$ and hence does not require normality of $\mathbf{W}_t$. The covariance of $\operatorname{vec}\mathbf{T}_{W-}^{*}$ is

$$\mathcal{E}\operatorname{vec}\mathbf{T}_{W-}^{*}(\operatorname{vec}\mathbf{T}_{W-}^{*})' = \boldsymbol{\Delta} \otimes \mathbf{I} \tag{5.14}$$

regardless of the nature of the distribution of $\mathbf{W}_t$.

The asymptotic covariance of $\operatorname{vec}\widehat{\boldsymbol{\Psi}}_k^*\boldsymbol{\Delta}$ and $\operatorname{vec}(\widehat{\boldsymbol{\Psi}}^* - \widehat{\boldsymbol{\Psi}}_k^*)\boldsymbol{\Delta}$ is

$$
\begin{aligned}
&A\operatorname{Cov}\left[\operatorname{vec}\widehat{\boldsymbol{\Psi}}_k^*\boldsymbol{\Delta}, \operatorname{vec}(\widehat{\boldsymbol{\Psi}}^* - \widehat{\boldsymbol{\Psi}}_k^*)\boldsymbol{\Delta}\right] \\
&\quad = \mathcal{E}\operatorname{vec}\left\{\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{T}_{W-}^{*21} & \mathbf{T}_{W-}^{*22} \end{bmatrix}\mathbf{M}\right\}\operatorname{vec}'\left\{\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{T}_{W-}^{*21} & \mathbf{T}_{W-}^{*22} \end{bmatrix}(\mathbf{I}-\mathbf{M})\right\} \\
&\quad = \mathbf{M}'\boldsymbol{\Delta}(\mathbf{I}-\mathbf{M}) \otimes \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \\
&\quad = \mathbf{0};
\end{aligned} \tag{5.15}
$$

that is, $\operatorname{vec}\widehat{\boldsymbol{\Psi}}_k^*$ and $\operatorname{vec}(\widehat{\boldsymbol{\Psi}}^* - \widehat{\boldsymbol{\Psi}}_k^*)$ are asymptotically independent. Hence, the covariance of the limiting distribution of $\operatorname{vec}\widehat{\boldsymbol{\Psi}}_k^*$ is the covariance of $\operatorname{vec}\widehat{\boldsymbol{\Psi}}^*$ minus the covariance of $\operatorname{vec}(\widehat{\boldsymbol{\Psi}}^* - \widehat{\boldsymbol{\Psi}}_k^*)$. The asymptotic covariance of $\operatorname{vec}(\widehat{\boldsymbol{\Psi}}^* - \widehat{\boldsymbol{\Psi}}_k^*)$ is

$$
\begin{aligned}
&A\operatorname{Cov}\left[\operatorname{vec}(\widehat{\boldsymbol{\Psi}}^* - \widehat{\boldsymbol{\Psi}}_k^*)\right] \\
&\quad = \boldsymbol{\Delta}^{-1}(\mathbf{I}-\mathbf{M}')\boldsymbol{\Delta}(\mathbf{I}-\mathbf{M})\boldsymbol{\Delta}^{-1} \otimes \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} = \boldsymbol{\Delta}^{-1}(\mathbf{I}-\mathbf{M}') \otimes \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \\
&\quad = (\boldsymbol{\Delta}^{-1} \otimes \mathbf{I})\left[(\mathbf{I}-\mathbf{M}') \otimes \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}\right].
\end{aligned} \tag{5.16}
$$

Each factor in the brackets is idempotent of rank $p-k$; that is, each can be diagonalized to $\operatorname{diag}(\mathbf{I}_{p-k}, \mathbf{0})$. In a sense the advantage of $\widehat{\boldsymbol{\Psi}}_k$ over $\widehat{\boldsymbol{\Psi}}$ is a reduction of the variability by a factor of $[(p-k)/p]^2$.

The covariance of the limiting distribution of $\operatorname{vec}\widehat{\mathbf{B}}_k^*$ is that of $\operatorname{vec}\widehat{\mathbf{B}}^*$ minus that of $\operatorname{vec}(\widehat{\mathbf{B}}^* - \widehat{\mathbf{B}}_k^*) = [\boldsymbol{\Phi} \otimes (\boldsymbol{\Phi}')^{-1}]\operatorname{vec}(\widehat{\boldsymbol{\Psi}}^* - \widehat{\boldsymbol{\Psi}}_k^*)$; that is, $(\boldsymbol{\Gamma}^{-1} \otimes \mathbf{I})$ minus

$$
\begin{aligned}
&A\operatorname{Cov}\left[\operatorname{vec}(\widehat{\mathbf{B}}^* - \widehat{\mathbf{B}}_k^*)\right] \\
&\quad = \left[\boldsymbol{\Phi} \otimes (\boldsymbol{\Phi}')^{-1}\right]A\operatorname{Cov}\left[\operatorname{vec}(\widehat{\boldsymbol{\Psi}}^* - \widehat{\boldsymbol{\Psi}}_k^*)\right](\boldsymbol{\Phi}' \otimes \boldsymbol{\Phi}^{-1}) \\
&\quad = \boldsymbol{\Phi}(\boldsymbol{\Delta}^{-1} - \boldsymbol{\Psi}_{1.}'\boldsymbol{\Theta}_1^{-1}\boldsymbol{\Psi}_{1.})\boldsymbol{\Phi}' \otimes (\boldsymbol{\Phi}')^{-1}\left[\mathbf{I} - \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\right]\boldsymbol{\Phi}^{-1}.
\end{aligned} \tag{5.17}
$$

Since $\mathbf{B}$ has rank k, it can be written as $\mathbf{B} = \boldsymbol{\Sigma}\boldsymbol{\Phi}_1\boldsymbol{\Phi}_1'\mathbf{B} = \boldsymbol{\Lambda}\boldsymbol{\Pi}'$, where $\boldsymbol{\Phi}_1$ consists of the first k columns of $\boldsymbol{\Phi}$, $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}\boldsymbol{\Phi}_1$ $(p \times p)$ and $\boldsymbol{\Pi} = \mathbf{B}'\boldsymbol{\Phi}_1 = \boldsymbol{\Phi}\boldsymbol{\Psi}_{1.}'$ $(p \times p)$. Note that $\boldsymbol{\Phi}\boldsymbol{\Delta}^{-1}\boldsymbol{\Phi}' = \boldsymbol{\Gamma}^{-1}$ since $\boldsymbol{\Phi}'\boldsymbol{\Gamma}\boldsymbol{\Phi} = \boldsymbol{\Delta}$. Further

$$
\boldsymbol{\Phi}\boldsymbol{\Psi}_{1.}'\boldsymbol{\Theta}_1^{-1}\boldsymbol{\Psi}_{1.}\boldsymbol{\Phi}' = \boldsymbol{\Pi}\boldsymbol{\Theta}_1^{-1}\boldsymbol{\Pi}', \tag{5.18}
$$

$\boldsymbol{\Theta}_1 = \boldsymbol{\Phi}_1'\mathbf{B}\boldsymbol{\Gamma}\mathbf{B}'\boldsymbol{\Phi}_1 = \boldsymbol{\Pi}'\boldsymbol{\Gamma}\boldsymbol{\Pi}$, $\mathbf{I} = \boldsymbol{\Phi}'\boldsymbol{\Sigma}\boldsymbol{\Phi}$, $(\boldsymbol{\Phi}')^{-1} = \boldsymbol{\Sigma}\boldsymbol{\Phi}$ and

$$
\boldsymbol{\Lambda}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda} = \boldsymbol{\Phi}_1'\boldsymbol{\Sigma}\boldsymbol{\Phi}_1 = \mathbf{I}_k. \tag{5.19}
$$

Then (5.17) is

$$
\begin{aligned}
&A\operatorname{Cov}\bigl[\operatorname{vec}(\widehat{\mathbf{B}}^{*}-\widehat{\mathbf{B}}_{k}^{*})\bigr]\\
&\quad=\bigl[\boldsymbol{\Gamma}^{-1}-\boldsymbol{\Pi}(\boldsymbol{\Pi}'\boldsymbol{\Gamma}\boldsymbol{\Pi})^{-1}\boldsymbol{\Pi}'\bigr]\otimes\bigl[\boldsymbol{\Sigma}-\boldsymbol{\Lambda}(\boldsymbol{\Lambda}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}'\bigr].
\end{aligned}
\tag{5.20}
$$

Note that the second term in each factor in (5.20) is invariant with respect to the transformation $(\boldsymbol{\Lambda},\boldsymbol{\Pi})\to(\boldsymbol{\Lambda}\mathbf{G},\boldsymbol{\Pi}\mathbf{G}^{-1})$ for nonsingular $\mathbf{G}$. When $\boldsymbol{\Lambda}=\boldsymbol{\Sigma}\boldsymbol{\Phi}_1$ and $\boldsymbol{\Pi}=\mathbf{B}'\boldsymbol{\Phi}_1$ $(\mathbf{G}=\mathbf{I})$, the effective normalization is $\boldsymbol{\Lambda}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}=\mathbf{I}$ and $\boldsymbol{\Pi}'\boldsymbol{\Gamma}\boldsymbol{\Pi}=\boldsymbol{\Theta}_1$.

THEOREM 4. *Let* $\mathbf{B}=\boldsymbol{\Lambda}\boldsymbol{\Pi}'$, *where* $\boldsymbol{\Lambda}$ *and* $\boldsymbol{\Pi}$ *are* $p\times k$ *matrices of rank* k. *Suppose that* $\{\mathbf{Z}_t\}$ *is a sequence of independent identically distributed random vectors with mean* $\mathbf{0}$ *and covariance* $\boldsymbol{\Sigma}$, *and let* $\{\mathbf{Y}_t\}$ *be defined by* $\mathbf{Y}_t=\mathbf{B}\mathbf{Y}_{t-1}+\mathbf{Z}_t$. *Then* $\sqrt{T}\operatorname{vec}(\widehat{\mathbf{B}}_k-\mathbf{B})$ *has a limiting normal distribution with mean* $\mathbf{0}$ *and covariance matrix*

$$
\boldsymbol{\Gamma}^{-1}\otimes\boldsymbol{\Sigma}-[\boldsymbol{\Gamma}^{-1}-\boldsymbol{\Pi}(\boldsymbol{\Pi}'\boldsymbol{\Gamma}\boldsymbol{\Pi})^{-1}\boldsymbol{\Pi}']\otimes[\boldsymbol{\Sigma}-\boldsymbol{\Lambda}(\boldsymbol{\Lambda}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}'].
\tag{5.21}
$$

This covariance matrix can be written as

$$
\begin{aligned}
&(\boldsymbol{\Gamma}^{-1}\otimes\boldsymbol{\Sigma})-(\boldsymbol{\Gamma}^{-1}\otimes\boldsymbol{\Sigma})\\
&\qquad\times\bigl\{\bigl[\mathbf{I}-\boldsymbol{\Gamma}\boldsymbol{\Pi}(\boldsymbol{\Pi}'\boldsymbol{\Gamma}\boldsymbol{\Pi})^{-1}\boldsymbol{\Pi}'\bigr]\otimes\bigl[\mathbf{I}-\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}(\boldsymbol{\Lambda}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}'\bigr]\bigr\}.
\end{aligned}
\tag{5.22}
$$

The first term in (5.22) is the covariance matrix of $\sqrt{T}\operatorname{vec}(\widehat{\mathbf{B}}-\mathbf{B})$, the least squares estimator of $\mathbf{B}$. The second term is the covariance matrix of the limiting distribution of $\sqrt{T}\operatorname{vec}(\widehat{\mathbf{B}}-\widehat{\mathbf{B}}_k)$. Each bracketed matrix in the pair of braces is idempotent of rank $p-k$.

A measure of the reduction in the variance of the estimator of $\mathbf{B}$ is

$$
\begin{aligned}
&\operatorname{tr}\bigl[(A\operatorname{Cov}\widehat{\mathbf{B}}^{*})^{-1}A\operatorname{Cov}(\widehat{\mathbf{B}}^{*}-\widehat{\mathbf{B}}_{k}^{*})\bigr]\big/\bigl[\operatorname{tr}(A\operatorname{Cov}\widehat{\mathbf{B}}^{*})^{-1}A\operatorname{Cov}\widehat{\mathbf{B}}^{*}\bigr]\\
&\quad=\operatorname{tr}\bigl[\mathbf{I}-\boldsymbol{\Gamma}\boldsymbol{\Pi}(\boldsymbol{\Pi}'\boldsymbol{\Gamma}\boldsymbol{\Pi})^{-1}\boldsymbol{\Pi}'\bigr]\operatorname{tr}\bigl[\mathbf{I}-\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}(\boldsymbol{\Lambda}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}'\bigr]/p^{2}\\
&\quad=(p-k)^{2}/p^{2}.
\end{aligned}
\tag{5.23}
$$

We have used $\operatorname{tr}(\mathbf{A}\otimes\mathbf{B})=\operatorname{tr}\mathbf{A}\operatorname{tr}\mathbf{B}$, $\operatorname{tr}\boldsymbol{\Gamma}\boldsymbol{\Pi}(\boldsymbol{\Pi}'\boldsymbol{\Gamma}\boldsymbol{\Pi})^{-1}\boldsymbol{\Pi}'=\operatorname{tr}\boldsymbol{\Pi}'\boldsymbol{\Gamma}\boldsymbol{\Pi}(\boldsymbol{\Pi}'\boldsymbol{\Gamma}\boldsymbol{\Pi})^{-1}=\operatorname{tr}\boldsymbol{\Theta}_1\boldsymbol{\Theta}_1^{-1}$, $\operatorname{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}(\boldsymbol{\Lambda}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}'=\operatorname{tr}\boldsymbol{\Lambda}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}(\boldsymbol{\Lambda}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda})^{-1}=\operatorname{tr}\mathbf{I}_k\mathbf{I}_k^{-1}=k$ and (5.19).

The limiting distribution of $\sqrt{T}(\widehat{\mathbf{B}}_k-\mathbf{B})$ holds under exactly the same conditions as for the limiting distribution of $\sqrt{T}(\widehat{\mathbf{B}}-\mathbf{B})$; in particular, only the second-order moment of $\mathbf{Z}_t$ is assumed. Hence confidence regions for $\mathbf{B}$ established on the basis of normality of $\mathbf{Z}_t$ hold generally. Note that the asymptotic distribution of the sample canonical variate coefficients depends on the fourth-order moments of $\mathbf{Z}_t$; nevertheless, the asymptotic distribution of the reduced rank regression estimator, which is a function of those coefficients, does not depend on fourth-order moments.

The asymptotic covariance matrix of $\widehat{\mathbf{B}}_k$ in this AR(1) model is the same as the asymptotic covariance matrix of $\widehat{\mathbf{B}}_k$ in the model $\mathbf{Y}_t=\mathbf{B}\mathbf{X}_t+\mathbf{Z}_t$, where $\mathbf{X}_t$'s

are independently identically distributed with $\mathcal{E}\mathbf{X}_t = \mathbf{0}$ and $\mathcal{E}\mathbf{X}_t\mathbf{X}_t' = \mathbf{\Sigma}_{XX}$ (in place of $\mathbf{\Gamma}$) or where the $\mathbf{X}_t$'s are nonstochastic and $\mathbf{S}_{XX} \to \mathbf{\Sigma}_{XX}$ [Anderson (1999b)]. In the present AR(1) model the transformation to canonical form $\mathbf{Y}_t \to \mathbf{\Phi}'\mathbf{Y}_t, \mathbf{Y}_{t-1} \to \mathbf{\Phi}'\mathbf{Y}_{t-1}$ is different from the transformation for independent $\mathbf{X}_t$'s ($\mathbf{Y}_t \to \mathbf{A}'\mathbf{Y}_t, \mathbf{X}_t \to \mathbf{\Omega}'\mathbf{X}_t$).

The likelihood ratio criterion for testing the null hypothesis that the rank of $\mathbf{B}$ is k against alternatives that the rank is greater than k when the $\mathbf{Z}_t$'s are normally distributed is

$$-2\log\lambda = -T\sum_{i=k+1}^{p}\log(1-r_i^2) \sim T\sum_{i=k+1}^{p} r_i^2 \tag{5.24}$$

[Anderson (1951)]. When the null hypothesis is true, this criterion has a limiting χ^2-distribution with $(p-k)^2$ degrees of freedom. This number is the product of the ranks of the two idempotent factors in (5.16) and (5.20).

6. Autoregressive processes of order greater than one.

6.1. *Canonical correlations and vectors.* We consider the process (2.1) written as

$$\mathbf{Y}_t = \mathbf{B}\widetilde{\mathbf{Y}}_{t-1} + \mathbf{Z}_t, \tag{6.1}$$

where $\mathbf{B} = (\mathbf{B}_1, \ldots, \mathbf{B}_m)$. Since $\mathbf{\Gamma} = \mathbf{B}\widetilde{\mathbf{\Gamma}}\mathbf{B}' + \mathbf{\Sigma}$, (4.1), which defines $\boldsymbol{\phi}$ and θ, is equivalent to (4.5); since $\mathbf{S}_{--} = \widehat{\mathbf{B}}\widetilde{\mathbf{S}}_{--}\widehat{\mathbf{B}}' + \mathbf{S}_{\widehat{Z}\widehat{Z}} + o_p(1/\sqrt{T})$ and $\mathbf{S}_{\widehat{Z}\widehat{Z}} = \mathbf{S}_{ZZ} + o_p(1/\sqrt{T})$, (4.3), defining $\mathbf{f}$ and t, is asymptotically equivalent to (4.6).

The transformation $\mathbf{X}_t = \mathbf{\Phi}'\mathbf{Y}_t$ leads to (4.16), which in turn leads to (4.17) and (4.18) for $\mathbf{H}^*$ and $\mathbf{D}^*$. To carry out the analysis, we need the asymptotic distribution of $\mathbf{S}_{--}$ and $\mathbf{S}_{ZZ}$ and of $\mathbf{T}_{--}$ and $\mathbf{T}_{WW}$ for $m > 1$.

The process $\{\widetilde{\mathbf{Y}}_t\}$ can be defined by

$$\widetilde{\mathbf{Y}}_t = \widetilde{\mathbf{B}}\widetilde{\mathbf{Y}}_{t-1} + \widetilde{\mathbf{Z}}_t, \tag{6.2}$$

where $\widetilde{\mathbf{Y}}_t = (\mathbf{Y}_t', \mathbf{Y}_{t-1}', \ldots, \mathbf{Y}_{t-m+1}')'$, $\widetilde{\mathbf{Z}}_t = (\mathbf{Z}_t', \mathbf{0}, \ldots, \mathbf{0})'$, and

$$\widetilde{\mathbf{B}} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 & \ldots & \mathbf{B}_{m-1} & \mathbf{B}_m \\ \mathbf{I} & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \ldots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{I} & \mathbf{0} \end{bmatrix} \tag{6.3}$$

[Anderson (1971), Section 5.3]. Note that the roots of (2.2) are the roots of $|\lambda\mathbf{I} - \widetilde{\mathbf{B}}| = 0$ (assumed to be less than 1 in absolute value). We have $\mathbf{Y}_t = \mathbf{J}\widetilde{\mathbf{Y}}_t$,

where $\mathbf{J} = (\mathbf{I}_p, \mathbf{0}, \ldots, \mathbf{0})$, $\mathbf{Z}_t = \mathbf{J}\widetilde{\mathbf{Z}}_t$, $\mathbf{S}_{YY} = \mathbf{J}\widetilde{\mathbf{S}}_{YY}\mathbf{J}'$ and $\mathbf{S}_{ZZ} = \mathbf{J}\widetilde{\mathbf{S}}_{ZZ}\mathbf{J}'$. The transformation $\mathbf{X}_t = \mathbf{\Phi}'\mathbf{Y}_t$, $\mathbf{W}_t = \mathbf{\Phi}'\mathbf{Z}_t$ carries (6.1) to

$$\mathbf{X}_t = \mathbf{\Psi}\widetilde{\mathbf{X}}_{t-1} + \mathbf{W}_t, \tag{6.4}$$

where $\mathbf{\Psi} = \mathbf{\Phi}'\mathbf{B}(\mathbf{I}_p \otimes \mathbf{\Phi}')^{-1}$. The process (6.2) is carried to $\widetilde{\mathbf{X}}_t = \widetilde{\mathbf{\Psi}}\widetilde{\mathbf{X}}_{t-1} + \widetilde{\mathbf{W}}_t$, where $\widetilde{\mathbf{\Psi}} = (\mathbf{I}_m \otimes \mathbf{\Phi}')\widetilde{\mathbf{B}}[\mathbf{I}_m \otimes (\mathbf{\Phi}')^{-1}]'$ has the same form as $\widetilde{\mathbf{B}}$. We have $\mathbf{X}_t = \mathbf{J}\widetilde{\mathbf{X}}_t$, $\mathbf{W}_t = \mathbf{J}\widetilde{\mathbf{W}}_t$, $\mathbf{T}_{XX} = \mathbf{J}\widetilde{\mathbf{T}}_{XX}\mathbf{J}'$, $\mathbf{T}_{WW} = \mathbf{J}\widetilde{\mathbf{T}}_{WW}\mathbf{J}'$. Let $\widetilde{\mathbf{\Delta}} = \mathcal{E}\widetilde{\mathbf{X}}_t\widetilde{\mathbf{X}}_t' = (\mathbf{I}_p \otimes \mathbf{\Phi}')\widetilde{\mathbf{\Gamma}}(\mathbf{I}_p \otimes \mathbf{\Phi})$. The asymptotic covariances of $\widetilde{\mathbf{T}}^*_{--} = \sqrt{T}(\widetilde{\mathbf{T}}_{--} - \widetilde{\mathbf{\Delta}})$ and $\widetilde{\mathbf{T}}^*_{WW} = \sqrt{T}(\widetilde{\mathbf{T}}_{WW} - \mathbf{I})$ are found from Theorem 1.

THEOREM 5. *If the $\mathbf{W}_t$ are independently normally distributed with $\mathcal{E}\mathbf{W}_t = \mathbf{0}$ and $\mathcal{E}\mathbf{W}_t\mathbf{W}_t' = \mathbf{I}$, then $\widetilde{\mathbf{T}}^*_{--}$ and $\widetilde{\mathbf{T}}^*_{WW}$ have a limiting normal distribution with means $\mathbf{0}$ and $\mathbf{0}$ and covariances*

$$\begin{aligned}\mathcal{E}\operatorname{vec}\widetilde{\mathbf{T}}^*_{--}(\operatorname{vec}\widetilde{\mathbf{T}}^*_{--})' &\to (\mathbf{I}+\widetilde{\mathbf{K}})\big[\mathbf{I}-(\widetilde{\mathbf{\Psi}}\otimes\widetilde{\mathbf{\Psi}})\big]^{-1}\\ &\quad\times\big[(\widetilde{\mathbf{\Delta}}\otimes\mathbf{I})+(\mathbf{I}\otimes\widetilde{\mathbf{\Delta}})-(\mathbf{I}\otimes\mathbf{I})\big]\big[\mathbf{I}-(\widetilde{\mathbf{\Psi}}'\otimes\widetilde{\mathbf{\Psi}}')\big]^{-1}\\ &= (\mathbf{I}+\widetilde{\mathbf{K}})\{\big[\mathbf{I}-(\widetilde{\mathbf{\Psi}}\otimes\widetilde{\mathbf{\Psi}})\big]^{-1}(\widetilde{\mathbf{\Delta}}\otimes\widetilde{\mathbf{\Delta}})\\ &\quad+(\widetilde{\mathbf{\Delta}}\otimes\widetilde{\mathbf{\Delta}})\big[\mathbf{I}-(\widetilde{\mathbf{\Psi}}'\otimes\widetilde{\mathbf{\Psi}})\big]^{-1}-(\widetilde{\mathbf{\Delta}}\otimes\widetilde{\mathbf{\Delta}})\},\end{aligned} \tag{6.5}$$

$$\mathcal{E}\operatorname{vec}\widetilde{\mathbf{T}}^*_{WW}(\operatorname{vec}\widetilde{\mathbf{T}}^*_{WW})' = (\mathbf{I}+\widetilde{\mathbf{K}})(\mathbf{J}'\mathbf{J}\otimes\mathbf{J}'\mathbf{J}) = (\mathbf{J}'\otimes\mathbf{J}')(\mathbf{I}+\mathbf{K})(\mathbf{J}\otimes\mathbf{J}), \tag{6.6}$$

$$\begin{aligned}\mathcal{E}\operatorname{vec}\widetilde{\mathbf{T}}^*_{--}(\operatorname{vec}\widetilde{\mathbf{T}}^*_{WW})' &\to (\mathbf{I}+\widetilde{\mathbf{K}})\big[\mathbf{I}-(\widetilde{\mathbf{\Psi}}\otimes\widetilde{\mathbf{\Psi}})\big]^{-1}(\mathbf{J}'\mathbf{J}\otimes\mathbf{J}'\mathbf{J})\\ &= \big[\mathbf{I}-(\widetilde{\mathbf{\Psi}}\otimes\widetilde{\mathbf{\Psi}})\big]^{-1}(\mathbf{J}'\otimes\mathbf{J}')(\mathbf{I}+\mathbf{K})(\mathbf{J}\otimes\mathbf{J}).\end{aligned} \tag{6.7}$$

Here $\widetilde{\mathbf{K}}$ refers to the commutation matrix of dimension $(pm)^2 \times (pm)^2$. Note $\widetilde{\mathbf{W}}_t = \mathbf{J}'\mathbf{W}_t$ and $\widetilde{\mathbf{K}}(\mathbf{J}'\otimes\mathbf{J}') = (\mathbf{J}'\otimes\mathbf{J}')\mathbf{K}$.

Since $\operatorname{vec}\mathbf{T}^*_{--} = (\mathbf{J}\otimes\mathbf{J})\operatorname{vec}\widetilde{\mathbf{T}}^*_{--}$ and $\operatorname{vec}\mathbf{T}^*_{WW} = (\mathbf{J}\otimes\mathbf{J})\operatorname{vec}\widetilde{\mathbf{T}}^*_{WW}$, the asymptotic covariances of $\operatorname{vec}\mathbf{T}^*_{--}$ are the asymptotic covariances given in Theorem 5 multiplied on the left by $(\mathbf{J}\otimes\mathbf{J})$ and the right by $(\mathbf{J}'\otimes\mathbf{J}')$. Define

$$\widetilde{\mathbf{\Lambda}} = \big[\mathbf{I}-(\widetilde{\mathbf{\Psi}}\otimes\widetilde{\mathbf{\Psi}})\big]^{-1}. \tag{6.8}$$

Then

$$\begin{aligned}\mathcal{E}\operatorname{vec}\mathbf{T}^*_{--}(\operatorname{vec}\mathbf{T}^*_{--})' &\to (\mathbf{J}\otimes\mathbf{J})(\mathbf{I}+\widetilde{\mathbf{K}})\{\widetilde{\mathbf{\Lambda}}(\widetilde{\mathbf{\Delta}}\otimes\widetilde{\mathbf{\Delta}})+(\widetilde{\mathbf{\Delta}}\otimes\widetilde{\mathbf{\Delta}})\widetilde{\mathbf{\Lambda}}'\}(\mathbf{J}'\otimes\mathbf{J}')\\ &\quad-(\mathbf{I}+\mathbf{K})(\mathbf{\Delta}\otimes\mathbf{\Delta})\\ &= (\mathbf{I}+\mathbf{K})(\mathbf{J}\otimes\mathbf{J})\widetilde{\mathbf{\Lambda}}(\widetilde{\mathbf{\Delta}}\otimes\widetilde{\mathbf{\Delta}})(\mathbf{J}'\otimes\mathbf{J}')\\ &\quad+(\mathbf{J}\otimes\mathbf{J})(\widetilde{\mathbf{\Delta}}\otimes\widetilde{\mathbf{\Delta}})\widetilde{\mathbf{\Lambda}}'(\mathbf{J}'\otimes\mathbf{J}')(\mathbf{I}+\mathbf{K})\\ &\quad-(\mathbf{I}+\mathbf{K})(\mathbf{\Delta}\otimes\mathbf{\Delta}),\end{aligned} \tag{6.9}$$

$$\mathcal{E}\operatorname{vec}\mathbf{T}^*_{WW}(\operatorname{vec}\mathbf{T}^*_{WW})' = (\mathbf{I}+\mathbf{K})(\mathbf{I}\otimes\mathbf{I}), \tag{6.10}$$

$$\mathcal{E}\operatorname{vec}\mathbf{T}^*_{--}(\operatorname{vec}\widetilde{\mathbf{T}}^*_{WW})' \to (\mathbf{I}+\mathbf{K})(\mathbf{J}\otimes\mathbf{J})\widetilde{\mathbf{\Lambda}}(\mathbf{J}'\otimes\mathbf{J}'). \tag{6.11}$$

The covariance of the limiting distribution of $\operatorname{vec}\mathbf{T}^*_{--} - (\mathbf{\Delta}\otimes\mathbf{I})\operatorname{vec}\mathbf{T}^*_{WW}$ is

$$\begin{aligned}
&\mathcal{E}\big[\operatorname{vec}\mathbf{T}^*_{--} - (\mathbf{\Delta}\otimes\mathbf{I})\operatorname{vec}\mathbf{T}^*_{WW}\big]\big[\operatorname{vec}\mathbf{T}^*_{--} - (\mathbf{\Delta}\otimes\mathbf{I})\operatorname{vec}\mathbf{T}^*_{WW}\big]' \\
&\quad= (\mathbf{I}+\mathbf{K})(\mathbf{J}\otimes\mathbf{J})\big[\widetilde{\mathbf{\Lambda}}(\widetilde{\mathbf{\Delta}}\otimes\widetilde{\mathbf{\Delta}}) + (\widetilde{\mathbf{\Delta}}\otimes\widetilde{\mathbf{\Delta}})\widetilde{\mathbf{\Lambda}}'\big](\mathbf{J}'\otimes\mathbf{J}') \\
&\quad\quad - (\mathbf{I}+\mathbf{K})(\mathbf{J}\otimes\mathbf{J})\widetilde{\mathbf{\Lambda}}(\mathbf{J}'\otimes\mathbf{J}')(\mathbf{\Delta}\otimes\mathbf{I}) \\
&\quad\quad - (\mathbf{\Delta}\otimes\mathbf{I})(\mathbf{J}\otimes\mathbf{J})\widetilde{\mathbf{\Lambda}}'(\mathbf{J}'\otimes\mathbf{J}')(\mathbf{I}+\mathbf{K}) \\
&\quad\quad - (\mathbf{I}+\mathbf{K})(\mathbf{\Delta}\otimes\mathbf{\Delta}) + (\mathbf{\Delta}\otimes\mathbf{I})(\mathbf{I}+\mathbf{K})(\mathbf{\Delta}\otimes\mathbf{I}).
\end{aligned} \tag{6.12}$$

The last line of (6.12) is $(\mathbf{\Delta}^2\otimes\mathbf{I}) - (\mathbf{\Delta}\otimes\mathbf{\Delta}) = (\mathbf{\Delta}\otimes\mathbf{I})\mathbf{N}$. Note that $(\mathbf{J}\otimes\mathbf{J})\widetilde{\mathbf{\Lambda}} \times(\mathbf{J}'\otimes\mathbf{J}')$ is the upper left-hand $p\times p$ submatrix of $\widetilde{\mathbf{\Lambda}} = [\mathbf{I} - (\widetilde{\mathbf{\Psi}}\otimes\widetilde{\mathbf{\Psi}})]^{-1} = \sum_{s=0}^{\infty}(\widetilde{\mathbf{\Psi}}^s\otimes\widetilde{\mathbf{\Psi}}^s)$ and $(\mathbf{J}\otimes\mathbf{J})\widetilde{\mathbf{\Lambda}}(\widetilde{\mathbf{\Delta}}\otimes\widetilde{\mathbf{\Delta}})(\mathbf{J}'\otimes\mathbf{J}')$ is the upper left-hand submatrix of

$$\begin{aligned}
\big[\mathbf{I} - (\widetilde{\mathbf{\Psi}}\otimes\widetilde{\mathbf{\Psi}})\big]^{-1}(\widetilde{\mathbf{\Delta}}\otimes\widetilde{\mathbf{\Delta}}) &= \sum_{s=0}^{\infty}(\widetilde{\mathbf{\Psi}}^s\otimes\widetilde{\mathbf{\Psi}}^s)(\widetilde{\mathbf{\Delta}}\otimes\widetilde{\mathbf{\Delta}}) \\
&= \sum_{s=0}^{\infty}(\mathcal{E}\widetilde{\mathbf{X}}_t\widetilde{\mathbf{X}}'_{t-s}\otimes\mathcal{E}\widetilde{\mathbf{X}}_t\widetilde{\mathbf{X}}'_{t-s}),
\end{aligned} \tag{6.13}$$

which is $\sum_{s=0}^{\infty}(\mathbf{M}_s\otimes\mathbf{M}_s)$, where $\mathbf{M}_s = \mathcal{E}\mathbf{X}_t\mathbf{X}'_{t-s}$. Let $\mathbf{J}\widetilde{\mathbf{\Psi}}^s\mathbf{J}' = \mathbf{\Pi}_s$. Then (6.12) can be written

$$\begin{aligned}
&A\operatorname{Cov}\big[\operatorname{vec}\mathbf{T}^*_{--} - (\mathbf{\Delta}\otimes\mathbf{I})\operatorname{vec}\mathbf{T}^*_{WW}, \operatorname{vec}\mathbf{T}^*_{--} - (\mathbf{\Delta}\otimes\mathbf{I})\operatorname{vec}\mathbf{T}^*_{WW}\big] \\
&\quad= (\mathbf{I}+\mathbf{K})\sum_{s=-\infty}^{\infty}(\mathbf{M}_s\otimes\mathbf{M}_s) - (\mathbf{I}+\mathbf{K})\sum_{s=0}^{\infty}(\mathbf{\Pi}_s\mathbf{\Delta}\otimes\mathbf{\Pi}_s) \\
&\quad\quad - \sum_{s=0}^{\infty}(\mathbf{\Delta}\mathbf{\Pi}'_s\otimes\mathbf{\Pi}'_s)(\mathbf{I}+\mathbf{K}) + (\mathbf{\Delta}^2\otimes\mathbf{I}) + \mathbf{K}(\mathbf{\Delta}\otimes\mathbf{\Delta}).
\end{aligned} \tag{6.14}$$

THEOREM 6. *If the $\mathbf{Z}_t$'s are independently normally distributed and if the roots of $|\mathbf{\Gamma} - \delta\mathbf{\Sigma}|$ are distinct, the nonzero elements of $\mathbf{D}^*$ have a limiting normal distribution with means* 0; *the covariance of this limiting normal distribution is*

$$\begin{aligned}
&A\operatorname{Cov}(\operatorname{vec}\mathbf{D}^*, \operatorname{vec}\mathbf{D}^*) \\
&\quad= 2\mathbf{E}(\mathbf{J}\otimes\mathbf{J})\big\{\widetilde{\mathbf{\Lambda}}(\widetilde{\mathbf{\Delta}}\otimes\widetilde{\mathbf{\Delta}}) + (\widetilde{\mathbf{\Delta}}\otimes\widetilde{\mathbf{\Delta}})\widetilde{\mathbf{\Lambda}}'\big\}(\mathbf{J}'\otimes\mathbf{J}')\mathbf{E} \\
&\quad\quad - 2\mathbf{E}\big\{(\mathbf{J}\otimes\mathbf{J})\widetilde{\mathbf{\Lambda}}(\mathbf{J}'\mathbf{\Delta}\otimes\mathbf{J}') + (\mathbf{\Delta}\mathbf{J}\otimes\mathbf{J})\widetilde{\mathbf{\Lambda}}'(\mathbf{J}'\otimes\mathbf{J}')\big\}\mathbf{E}.
\end{aligned} \tag{6.15}$$

In the calculations of (6.15) $\widetilde{\mathbf{\Delta}}$ can be found from $\widetilde{\mathbf{\Delta}} = \widetilde{\mathbf{\Psi}}\widetilde{\mathbf{\Delta}}\widetilde{\mathbf{\Psi}}' + \mathbf{I}$. The operation $(\boldsymbol{\varepsilon}'_i\otimes\boldsymbol{\varepsilon}'_i)(\mathbf{J}\otimes\mathbf{J})\ \{\cdots\}(\mathbf{J}'\otimes\mathbf{J}')(\boldsymbol{\varepsilon}_j\otimes\boldsymbol{\varepsilon}_j)$ selects the element in the i,ith row and j,jth column of $\{\cdots\}$, and $(\boldsymbol{\varepsilon}_i\otimes\boldsymbol{\varepsilon}_i)(\boldsymbol{\varepsilon}'_i\otimes\boldsymbol{\varepsilon}'_i)(\mathbf{J}\otimes\mathbf{J})\{\cdots\}(\mathbf{J}'\otimes\mathbf{J}')(\boldsymbol{\varepsilon}_j\otimes\boldsymbol{\varepsilon}_j)(\boldsymbol{\varepsilon}'_j\otimes\boldsymbol{\varepsilon}'_j)$ places it in the i,ith row and j,jth column of the product. Since $\mathbf{\Delta}$ is a diagonal matrix, $(\boldsymbol{\varepsilon}'_i\otimes\boldsymbol{\varepsilon}_i)(\mathbf{J}\otimes\mathbf{J})\widetilde{\mathbf{\Lambda}}(\mathbf{J}'\mathbf{\Delta}\otimes\mathbf{J}')(\boldsymbol{\varepsilon}_j\otimes\boldsymbol{\varepsilon}_j)$ selects the i,ith row and the j,jth column and multiplies it by δ_j.

6.2. *Reduced rank regression.* The reduced rank regression estimator of $\mathbf{B}$ is (5.1) with $\widehat{\mathbf{B}}$ defined in (2.10) and $\mathbf{F}_1$ the first k columns of $\mathbf{F}$. The transformation $\mathbf{X}_t = \mathbf{\Phi}'\mathbf{Y}_t$, $\mathbf{W}_t = \mathbf{\Phi}'\mathbf{Z}_t$ carries (6.1) to (6.4). The reduced rank regression estimator of $\mathbf{\Psi}$ is (5.2), where

$$\widehat{\mathbf{\Psi}} = \mathbf{T}_{X-}\widetilde{\mathbf{T}}_{--}^{-1} = \mathbf{\Psi} + \mathbf{T}_{W-}\widetilde{\mathbf{T}}_{--}^{-1} \tag{6.16}$$

is the least squares estimator of $\mathbf{\Psi}$. From $\mathbf{\Delta} - \mathbf{I} = \mathbf{\Psi}\widetilde{\mathbf{\Delta}}\mathbf{\Psi}'$ we obtain $\mathbf{0} = \mathbf{\Psi}_{2\cdot}\widetilde{\mathbf{\Delta}}\mathbf{\Psi}_{2\cdot}'$, which implies $\mathbf{\Psi}_{2\cdot} = \mathbf{0}$; here $\mathbf{\Psi}'$ has been partitioned into k and $p-k$ columns, $\mathbf{\Psi}' = (\mathbf{\Psi}_{1\cdot}', \mathbf{\Psi}_{2\cdot}')$. The analog of (5.6) is

$$\widehat{\mathbf{\Psi}}_k^* = \begin{bmatrix} (\mathbf{T}_{WW}^{*11} + \mathbf{H}_{11}^* + \mathbf{H}_{11}^{*\prime})\mathbf{\Psi}_{1\cdot} \\ (\mathbf{T}_{WW}^{*21} + \mathbf{H}_{21}^*)\mathbf{\Psi}_{1\cdot} \end{bmatrix} + \begin{bmatrix} \mathbf{T}_{W-}^{*1\cdot} \\ \mathbf{0} \end{bmatrix} \widetilde{\mathbf{\Delta}}^{-1} + o_p(1). \tag{6.17}$$

Here $\mathbf{T}_{W-}^{*1\cdot}$ denotes the first k rows of $\mathbf{T}_{W-}^*$.

From (5.7) we have

$$\mathbf{H}_{21}^* = (\mathbf{T}_{XX}^{*21} - \mathbf{T}_{WW}^{*21}\mathbf{\Delta}_1)(\mathbf{\Delta}_1 - \mathbf{I})^{-1} + o_p(1). \tag{6.18}$$

From (6.4) and $\mathbf{\Psi}_{2\cdot} = \mathbf{0}$ we obtain

$$\mathbf{T}_{XX}^{21} = \mathbf{T}_{W-}\mathbf{\Psi}_{1\cdot}' + \mathbf{T}_{WW}^{21} + o_p(1), \tag{6.19}$$

$$\mathbf{H}_{21}^* = \mathbf{T}_{W-}^*\mathbf{\Psi}_{1\cdot}'(\mathbf{\Delta}_1 - \mathbf{I})^{-1} - \mathbf{T}_{WW}^{*21} + o_p(1). \tag{6.20}$$

Then

$$\begin{aligned} \widehat{\mathbf{\Psi}}_k^* &= \begin{bmatrix} \mathbf{0} \\ \mathbf{T}_{W-}^*\mathbf{\Psi}_{1\cdot}'(\mathbf{\Delta}_1 - \mathbf{I})^{-1}\mathbf{\Psi}_{1\cdot} \end{bmatrix} + \begin{bmatrix} \mathbf{T}_{W-}^{*1\cdot} \\ \mathbf{0} \end{bmatrix} \widetilde{\mathbf{\Delta}}^{-1} + o_p(1) \\ &= \left\{ \begin{bmatrix} \mathbf{T}_{W-}^{*1\cdot} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{T}_{W-}^{*2\cdot} \end{bmatrix} \widetilde{\mathbf{M}} \right\} \widetilde{\mathbf{\Delta}}^{-1} + o_p(1), \end{aligned} \tag{6.21}$$

where

$$\widetilde{\mathbf{M}} = \mathbf{\Psi}_{1\cdot}'\mathbf{\Theta}_1^{-1}\mathbf{\Psi}_{1\cdot}\widetilde{\mathbf{\Delta}}. \tag{6.22}$$

Note that $\widetilde{\mathbf{M}}^2 = \widetilde{\mathbf{M}}$. For (5.13) we obtain

$$\operatorname{vec}\big[(\widehat{\mathbf{\Psi}}^* - \widehat{\mathbf{\Psi}}_k^*)\widetilde{\mathbf{\Delta}}\big] = \big[(\mathbf{I} - \widetilde{\mathbf{M}}') \otimes \mathbf{I}\big] \operatorname{vec} \begin{bmatrix} \mathbf{0} \\ \mathbf{T}_{W-}^{*2\cdot} \end{bmatrix} + o_p(1), \tag{6.23}$$

and $\operatorname{vec}(\widehat{\mathbf{\Psi}}^* - \widehat{\mathbf{\Psi}}_k^*)$ and $\operatorname{vec}\widehat{\mathbf{\Psi}}_k^*$ are asymptotically independent. The covariance of the limiting distribution of $\operatorname{vec}(\widehat{\mathbf{\Psi}}^* - \widehat{\mathbf{\Psi}}_k^*)$ is (5.16) with $\mathbf{\Delta}$ replaced by $\widetilde{\mathbf{\Delta}}$ and $\mathbf{\Delta}^{-1}\mathbf{M}'$ replaced by $\mathbf{\Psi}_{1\cdot}'\mathbf{\Theta}_1\mathbf{\Psi}_{1\cdot} = \widetilde{\mathbf{\Delta}}^{-1}\widetilde{\mathbf{M}}'$. By algebra similar to that of Section 5 the covariance of the limiting distribution of $\sqrt{T}\operatorname{vec}(\widehat{\mathbf{B}}_r - \mathbf{B})$ is (5.20) with $\mathbf{\Gamma}$ and $\mathbf{\Pi}$ replaced by $\widetilde{\mathbf{\Gamma}}$ and $\widetilde{\mathbf{\Pi}}$, respectively, where $\widetilde{\mathbf{\Pi}}$ is defined by $\mathbf{B} = \mathbf{\Lambda}\widetilde{\mathbf{\Pi}}'$.

7. Discussion.

7.1. *Reinsel.* Velu, Reinsel and Wichern (1986), Ahn and Reinsel (1988), Reinsel (1997) and Reinsel and Velu (1998) treat the reduced rank regression estimator for autoregressive processes.

In the first paper the authors suggest the model in which $\mathbf{B}$ $(p \times pm)$ is factored into the product of a $p \times r$ matrix and an $r \times pm$ matrix and proceed to find the asymptotic distribution of the two factors when $\mathbf{\Sigma}$ is assumed known. However, this distribution is incorrect because the variability due to $\mathbf{S}_{--}$ (my notation) is ignored. Further, the assertion that this asymptotic distribution holds when $\mathbf{\Sigma}$ is replaced by $\mathbf{S}_{ZZ}$ is incorrect. To explain this matter in more detail, suppose $\mathbf{\Sigma} = \mathbf{I}$ and $m = 1$. Then the first factor in $\mathbf{B} = \mathbf{\Lambda}\mathbf{\Pi}'$ is $\mathbf{\Lambda} = \mathbf{\Phi}_1 = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_k)$ as defined in (4.1) with $\mathbf{\Sigma} = \mathbf{I}$, and its estimator is $\mathbf{F}_1 = (\mathbf{f}_1, \ldots, \mathbf{f}_k)$ as defined in (4.3) with $\mathbf{S}_{ZZ}$ replaced by $\mathbf{I}$. The left-hand side equation in (4.3) leads to

$$
\begin{aligned}
&(\mathbf{B}\mathbf{S}^*_{--}\mathbf{B}' + \mathbf{S}^*_{Z-}\mathbf{B}' + \mathbf{B}\mathbf{S}^*_{-Z})\mathbf{\Phi}_1 \\
&\qquad = \mathbf{\Phi}_1\mathbf{T}^*_1 + \mathbf{F}^*_1\mathbf{\Theta}_1 - \mathbf{B}\mathbf{\Gamma}\mathbf{B}'\mathbf{F}^*_1 + o_p(1),
\end{aligned} \tag{7.1}
$$

where $\mathbf{F}^*_1 = \sqrt{T}(\mathbf{F}_1 - \mathbf{\Phi}_1)$ and $\mathbf{T}^*_1 = \sqrt{T}(\mathbf{T}_1 - \mathbf{\Theta}_1)$. After the transformation to $\mathbf{X}_t = \mathbf{\Phi}'\mathbf{Y}_t$, $\mathbf{W}_t = \mathbf{\Phi}'\mathbf{Z}_t$ (7.1) is

$$
\begin{aligned}
&(\mathbf{\Psi}\mathbf{T}^*_{--}\mathbf{\Psi}' + \mathbf{T}^*_{W-}\mathbf{\Psi}' + \mathbf{\Psi}\mathbf{T}^*_{-W})\mathbf{I}_{(k)} \\
&\qquad = \mathbf{I}_{(k)}\mathbf{T}^*_1 + \mathbf{H}^*_1\mathbf{\Theta}_1 - \mathbf{\Theta}_1\mathbf{H}^*_1 + o_p(1).
\end{aligned} \tag{7.2}
$$

In solving for $\mathbf{H}^*_1$ the term $\mathbf{\Psi}\mathbf{T}^*_{--}\mathbf{\Psi}'$ was discarded by the authors.

Ahn and Reinsel (1988) generalize the model to let $\mathbf{B}_1 = \mathbf{B}_{1L}\mathbf{B}_{1R}, \ldots, \mathbf{B}_m = \mathbf{B}_{mL}\mathbf{B}_{mR}$ such that range $\mathbf{B}_{iL} \supset$ range $B_{i+1,L}$ and find a canonical form for the matrices. They set up the likelihood equations and indicate that the covariance matrix of the asymptotic normal distribution of the estimators can be obtained from the Fisher information matrix. When their results are specialized to the first-order case ($m = 1$), their results agree with Theorem 4, but their expression for the asymptotic covariance matrix is not explicit (personal communication by one of the authors). See also Reinsel (1997), Section 6.1, for further details on the nested model.

In Reinsel and Velu (1998) the Theorem 2 of Velu, Reinsel and Wichern (1986) about the asymptotic distribution of the factors and its proof are repeated as Theorem 2.4 and its proof as pertaining to the model $\mathbf{Y}_t = \mathbf{B}\mathbf{X}_t + \mathbf{Z}_t$ with $\mathcal{E}\mathbf{Z}_t\mathbf{Z}'_t = \mathbf{\Sigma}_{ZZ}$ known and $\mathbf{X}_t$ being exogenous or independent regressors. In this case the variability due to $\mathbf{S}_{XX} - \mathbf{\Sigma}_{XX}$ is ignored. In the section on the autoregressive model they approach the asymptotic distribution of the reduced rank regression estimator by assuming that the $\mathbf{Z}_t$'s are normally distributed and use the Fisher information matrix. Although they do not give the asymptotic covariance matrix explicitly, the details can be filled in, as shown by one of the authors in a personal communication; the asymptotic covariance matrix is shown to be (5.21).

7.2. *Lütkepohl.* Lütkepohl (1993) purports to obtain the asymptotic distribution of $\widehat{\mathbf{B}}_k = \widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Pi}}'$ by finding the joint asymptotic distribution of $\widehat{\mathbf{\Lambda}}$ and $\widehat{\mathbf{\Pi}}$ and from that the distribution of $\widehat{\mathbf{\Lambda}}\mathbf{\Pi}' + \mathbf{\Lambda}\widehat{\mathbf{\Pi}}$. The asymptotic distribution of $\widehat{\mathbf{\Lambda}}$ and $\widehat{\mathbf{\Pi}}$ is incorrect, however; see Anderson (1999b). The asserted asymptotic covariance of $\widehat{\mathbf{B}}_k$ is not given very explicitly.

Acknowledgment. The author thanks two anonymous referees for valuable suggestions.

REFERENCES

AHN, S. K. and REINSEL, G. C. (1988). Nested reduced-rank autoregressive models for multiple time series. *J. Amer. Statist. Assoc.* **83** 849–856.

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

ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York.

ANDERSON, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. Wiley, New York.

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

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

BOX, G. E. P. and TIAO, G. C. (1977). A canonical analysis of multiple time series. *Biometrika* **64** 355–365.

LÜTKEPOHL, H. (1993). *Introduction to Multiple Time Series Analysis*, 2nd ed. Springer, New York.

REINSEL, G. C. (1997). *Elements of Multivariate Time Series Analysis*, 2nd ed. Springer, New York.

REINSEL, G. C. and VELU, R. P. (1998). *Multivariate Reduced Rank Regression*. Springer, New York.

RYAN, D. A. J., HUBERT, J. J., CARTER, E. M., SPRAGUE, J. B. and PARROTT, J. (1992). A reduced-rank multivariate regression approach to aquatic joint toxicity experiments. *Biometrics* **48** 155–162.

SCHMIDLI, H. (1995). *Reduced-Rank Regression*: *With Applications to Quantitative Structure-Activity Relationships*. Springer, Berlin.

STOICA, P. and VIBERG, M. (1996). Maximum likelihood parameter and rank estimation in reduced-rank multivariate linear regressions. *IEEE Trans. Signal Process.* **44** 3069–3078.

TIAO, G. C. and TSAY, R. S. (1989). Model specification in multivariate time series (with discussion). *J. Roy. Statist. Soc. Ser. B* **51** 157–213.

VELU, R. P., REINSEL, G. C. and WICHERN, D. W. (1986). Reduced rank models for multiple times series. *Biometrika* **73** 109–118.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-4065
E-MAIL: twa@stat.stanford.edu