Order determination in general vector autoregressions

Bent Nielsen

Department of Economics, University of Oxford
& Nuffield College, Oxford OX1 1NF, UK
E-mail: `bent.nielsen@nuf.ox.ac.uk`
`http://www.nuff.ox.ac.uk/users/nielsen`

12 July 2001

Abstract: In the application of autoregressive models the order of the model is often estimated using either a sequence of likelihood ratio tests or a likelihood based information criterion. The consistency of such procedures has been discussed extensively under the assumption that the characteristic roots of the autoregression are stationary. It is shown that these methods can be used regardless of the assumption to the characteristic roots.

1 Introduction

Order determination for stationary autoregressive time series has been discussed extensively in the literature. The two prevailing methods are either to test whether the last lag is redundant using a likelihood based test based on the χ^2-distribution or to estimate the lag length consistently using an information criteria. It is shown that these methods can be used regardless of the assumption of stationarity.

The statistical model is given by a p-dimensional time series (X_t) satisfying a kth order vector autoregressive equation

$$X_t = \sum_{l=1}^{k} A_l X_{t-l} + \mu D_t + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{1.1}$$

Here the innitial values $X_0, \ldots, X_{1-k}$ are fixed, the component D_t is a vector of deterministic terms such as a constant, a linear trend, or seasonal dummies, while the innovations, (ε_t), are assumed to be independently, identically normal, $\mathsf{N}_p(0, \Omega)$, distributed.

The aim is to determine the largest non-trivial order for the time series, k_0 say, so $A_{k_0} \neq 0$ and $A_j = 0$ for $j > k_0$. Two approaches are available of which the first is based on a likelihood ratio test for $A_k = 0$. The log likelihood ratio test statistic is

$$LR\,(A_k = 0) = T \log\det\hat{\Omega}_{k-1} - T \log\det\hat{\Omega}_k,$$

and this will be proved to be asymptotically χ^2. The second approach is to estimate k_0 by minimising the penalised likelihood

$$\Phi_j = \log\det\hat{\Omega}_j + j\frac{f(T)}{T}, \qquad j = 0,\ldots,k. \tag{1.2}$$

In the literature there are several candidates for the penalty function f. Akaike has $f(T) = 2p^2$, Schwarz (1978) has $f(T) = p^2 \log T$ while Hannan and Quinn (1979) and Quinn (1980) have $f(T) = 2p^2 \log\log T$. For stationary processes without deterministic components it has been shown that the estimator $\hat{k}_0$ is weakly consistent if $f(T) = \mathrm{o}(T)$ and $f(T) \to \infty$ as T increases, while Hannan and Quinn show that strong consistency is obtained if and only if $\limsup_{T\to\infty} f(T) > 2$. In other words the estimators of Hannan and Quinn and of Schwarz are consistent whiles Akaike's estimator is inconsistent. Some generalisations to non-stationary processes have been given by for instance Paulsen (1984), Pötscher (1989) and Tsay (1984). In the following some further generalisations are made with a view towards abandoning the stationarity condition altogether and including deterministic components.

The following notation is used throughout the paper: For a matrix α let $\alpha^{\otimes 2} = \alpha\alpha'$ and let $\|\alpha\|$ be the Euclidean norm. When α is symmetric then $\lambda_{\min}(\alpha)$ and $\lambda_{\max}(\alpha)$ denote the smallest and the largest eigenvalue respectively. While $\mathsf{E}(\varepsilon_t|\mathcal{F}_{t-1})$ is a conditional expectation the notation $(Y_t|Z_t)$ denotes the residual of the least squares regression of Y_t on Z_t. The abbreviations $a.s.$ and P are used for properties holding almost surely and in probability, respectively.

2 Results

The asymptotic analysis is to a large extent based on results of Lai and Wei (1985) with appropriate modifications presented by Nielsen (2001). Following the precedence of Lai and Wei the assumptions to the innovation process can be relaxed so the sequence of innovations (ε_t) is a martingale difference sequence with respect to an increasing sequence of σ-fields $(\mathcal{F}_t)$, that is ε_t is $\mathcal{F}_t$-measurable with $\mathsf{E}(\varepsilon_t|\mathcal{F}_{t-1}) = 0$ $a.s.$ and is assumed to satisfy the conditions

$$\sup_t \mathsf{E}\left(\|\varepsilon_t\|^{2+\gamma}|\mathcal{F}_{t-1}\right) \overset{a.s.}{<} \infty \qquad \text{for some } \gamma > 2, \tag{2.1}$$

$$\liminf_{T\to\infty} \lambda_{\min}\mathsf{E}(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}) \overset{a.s.}{>} 0. \tag{2.2}$$

The deterministic process will be assumed to satisfy the assumption

$$D_t = \mathbf{D}D_{t-1}, \qquad |\mathrm{eigen}(\mathbf{D})| = 1, \qquad \mathrm{rank}(D_1,\ldots,D_{\dim\mathbf{D}}) = \dim\mathbf{D}. \tag{2.3}$$

This condition ensures that the deterministic regressors are not collinear and is inspired by Johansen (2000).

In the analysis it is convenient to introduce the companion form

$$\begin{pmatrix} \mathbf{X}_t \\ D_t \end{pmatrix} = \begin{pmatrix} \mathbf{B} & \boldsymbol{\mu} \\ 0 & \mathbf{D} \end{pmatrix} \begin{pmatrix} \mathbf{X}_{t-1} \\ D_{t-1} \end{pmatrix} + \begin{pmatrix} \mathbf{e}_t \\ 0 \end{pmatrix}, \tag{2.4}$$

where $\mathbf{X}_t = (X'_t, \ldots, X'_{t-k+1})'$ and

$$\mathbf{B} = \left\{ \begin{array}{cc} A_1 \cdots A_{k-1} & A_k \\ I_{p(k-1)} & 0 \end{array} \right\}, \qquad \boldsymbol{\iota} = \left\{ \begin{array}{c} I_p \\ 0_{(k-1)p \times p} \end{array} \right\}, \qquad \boldsymbol{\mu} = \boldsymbol{\iota}\mu\mathbf{D}, \qquad \mathbf{e}_t = \boldsymbol{\iota}\varepsilon_t.$$

The process $\mathbf{X}$ can be decomposed using a similarity transformation. Following Herstein (1975, p. 308) there exists a regular, real matrix M which transforms $\mathbf{B}$ into a real, rational canonical form. In particular, M can be chosen so $M\mathbf{B}M^{-1} = \operatorname{diag}(\mathbf{U}, \mathbf{V}, \mathbf{W})$ is a block diagonal matrix where the absolute values of the eigenvalues of $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ are smaller than one, equal to one and at least one, respectively. The process $\mathbf{X}_t$ can therefore be decomposed as

$$M\mathbf{X}_t = \begin{pmatrix} U_t \\ V_t \\ W_t \end{pmatrix} = \begin{pmatrix} \mathbf{U} & 0 & 0 & \mu_U \\ 0 & \mathbf{V} & 0 & \mu_V \\ 0 & 0 & \mathbf{W} & \mu_W \end{pmatrix} \begin{pmatrix} U_{t-1} \\ V_{t-1} \\ W_{t-1} \\ D_t \end{pmatrix} + \begin{pmatrix} e_{U,t} \\ e_{V,t} \\ e_{W,t} \end{pmatrix}.$$

Finally, there exists a constant $\tilde{\mu}_U$, see Nielsen (2001, Lemma 2.1), so

$$U_t = \tilde{U}_t + \tilde{\mu}_U D_t \qquad \text{where} \qquad \tilde{U}_t = \mathbf{U}\tilde{U}_{t-1} + e_{U,t}. \tag{2.5}$$

The likelihood ratio test statistic is known to be asymptotically χ^2 as long as $|\text{eigen}(\mathbf{B})| < 1$, see Lütkepohl (1991, Section 4.2.2), but the result holds regardless of the assumption to $\mathbf{B}$. In order to use a Central Limit Theorem for the martingale difference sequence $\tilde{U}_{t-1}\varepsilon_t$ it is necessary to assume that the innovations have fourth moments and that

$$\mathsf{E}(\varepsilon_t^{\otimes 2} | \mathcal{F}_{t-1}) \overset{a.s.}{=} \Omega. \tag{2.6}$$

It can be proved along the lines of Chan and Wei (1988) that condition (2.6) ensures a more technical condition

$$\mathsf{P}\left(T^{-\eta-1} \sum_{t=1}^{T} V_t^{\otimes 2} > 0 \right) \to 1, \tag{2.7}$$

for $\eta = 1$. In the proofs it actually suffices that (2.7) holds for some $\eta > 0$.

Theorem 2.1 *Suppose Assumptions* (2.1), (2.2), (2.3), (2.6) *are satisfied. Then LR is asymptotically* $\chi^2(p^2)$.

Turning to the information criteria Paulsen (1984) and Tsay (1984) prove weak consistency for the case $|\text{eigen}(\mathbf{B})| \leq 1$. Once again the assumption to $\mathbf{B}$ is redundant.

Theorem 2.2 *Suppose Assumptions* (2.1), (2.2), (2.3) *and* (2.6) *are satisfied. If* $f(T) = \mathrm{o}(T)$ *and* $f(T) \to \infty$ *then* $\hat{k}_0 \xrightarrow{\mathsf{P}} k_0$.

Strong consistency is harder to prove. In the case where $|\text{eigen}(\mathbf{B})| < 1$ Hannan and Quinn (1979) and Quinn (1980) use a Law of the Iterated Logarithm to show a necessary and sufficient condition for consistency. While I believe the assumption to $\mathbf{B}$ is redundant that degree of generalisation has not quite been achieved. The problem is to prove an almost sure version of (2.7) such as

$$\liminf_{T \to \infty} \left(T^{-\eta-1} \sum_{t=1}^{T} V_t^{\otimes 2} \right) \overset{a.s.}{>} 0, \tag{2.8}$$

for some $\eta > 0$. This is essentially proved for $\eta = 0$ by Lai and Wei (1985, Theorem 3), but in view of (2.7) it is not unreasonable that it should hold for an $\eta > 0$. For the case where $\mathbf{V}$ has distinct eigenvalues this can be proved using Donsker and Varadhan's (1977) Law of the Iterated Logarithm for the integrated squared Brownian motion,

$$\liminf_{T \to \infty} \frac{\log \log T}{T^2} \int_0^T B_u^2 du \overset{a.s.}{=} \frac{1}{4},$$

in conjunction with a Skorokhod embedding.

The first of the following three generalisations of Hannan and Quinn's result avoids the condition (2.8) at the cost of a strong condition to the penalty function. Pötscher (1989, Theorem 3.1) proved the result for $|\text{eigen}(\mathbf{B})| \leq 1$, in which case it actually suffices that $\gamma > 0$ in Assumption (2.1), see also Lemma 3.2 below.

Theorem 2.3 *Suppose Assumptions* (2.1), (2.2), (2.3) *are satisfied. If* $f(T) = \mathrm{o}(T)$ *and* $f(T)/\log T \to \infty$ *then* $\hat{k}_0 \to k_0$ *a.s.*

Weaker conditions to the penalty function are sufficient under the condition (2.8). This result was proved first by Pötscher (1989, Theorem 3.4) for $|\text{eigen}(\mathbf{B})| < 1$.

Theorem 2.4 *Suppose Assumptions* (2.1), (2.2), (2.3), (2.8) *are satisfied. If* $f(T) = \mathrm{o}(T)$ *and* $f(T)/\log \log T \to \infty$ *then* $\hat{k}_0 \to k_0$ *a.s.*

As in Hannan and Quinn (1979) a more precise statement can be made under the additional assumption of stationarity and ergodicity of the innovations.

Theorem 2.5 *Suppose Assumptions* (2.1), (2.2), (2.3), (2.8) *are satisfied and that* (ε_t) *is stationary and ergodic. If* $f(T) = \mathrm{o}(T)$ *then it holds that* $\hat{k}_0 \to k_0$ *a.s. if and only if* $\limsup_{T \to \infty} (2 \log \log T)^{-1} f(T) > 1$ *a.s.*

3 Proofs

The likelihood ratio test statistic can be expressed in terms of the sample partial correlation of the residuals of X_t and X_{t-k} corrected for the lags of the process $\mathbf{X}_{t-1}$ and the deterministic components D_t. The main idea of the proof is to show that this partial correlation is equivalent to the correlation of the innovation ε_t and a component which is derived from the zero mean stationary component $\tilde{U}_{t-1}$. This idea is developed in a first subsection and subsequently the main theorems are proved along the lines of Hannan and Quinn (1979).

3.1 *Partial correlations*

Two expressions for the sample partial correlation are found in Lemmas 3.1, 3.5 with slightly stronger assumptions to the innovation process in the latter result. Both expressions involve a remainder term

$$R_T = \sum_{t=1}^{T} \varepsilon_t V_{t-2}' \left(\sum_{t=1}^{T} V_{t-2}^{\otimes 2} \right)^{-1} \sum_{t=1}^{T} V_{t-2} \varepsilon_t' - \sum_{t=1}^{T} \varepsilon_t V_{t-1}' \left(\sum_{t=1}^{T} V_{t-1}^{\otimes 2} \right)^{-1} \sum_{t=1}^{T} V_{t-1} \varepsilon_t',$$

which is discussed further in Lemma 3.6. Throughout this section it is assumed that equation (1.1) holds with $A_k = 0$ for some $k > 1$.

Lemma 3.1 *Suppose Assumptions* (2.1), (2.2), (2.3) *are satisfied. Then*

$$\begin{aligned} T S_{01} S_{11}^{-1} S_{10} \overset{a.s.}{=} & \sum_{t=1}^{T} \varepsilon_t \tilde{U}_{t-2}' \left(\sum_{t=1}^{T} \tilde{U}_{t-2}^{\otimes 2} \right)^{-1} \sum_{t=1}^{T} \tilde{U}_{t-2} \varepsilon_t' - \sum_{t=1}^{T} \varepsilon_t \tilde{U}_{t-1}' \left(\sum_{t=1}^{T} \tilde{U}_{t-1}^{\otimes 2} \right)^{-1} \sum_{t=1}^{T} \tilde{U}_{t-1} \varepsilon_t' \\ & + \sum_{t=1}^{T} \varepsilon_t \varepsilon_{t-1}' \left(\sum_{t=1}^{T} \varepsilon_{t-1}^{\otimes 2} \right)^{-1} \sum_{t=1}^{T} \varepsilon_{t-1} \varepsilon_t' + \mathrm{o}(1) + R_T. \end{aligned}$$

For the asymptotic analysis a series of results by Nielsen (2001) are needed. This work is to a large extent based on Lai and Wei (1985) as well as Wei (1985, Lemma 2).

Lemma 3.2 *Suppose Assumptions* (2.1), (2.2), (2.3) *are satisfied.*
If (Y_t, Z_t) *is either of the pairs* (U_t, V_t), (U_t, W_t), (V_t, W_t) *then*

$$\left\{ \sum_{t=1}^{T} (Y_t | D_t)^{\otimes 2} \right\}^{-1/2} \sum_{t=1}^{T} (Y_t | D_t) (Z_t | D_t)' \left\{ \sum_{t=1}^{T} (Z_t | D_t)^{\otimes 2} \right\}^{-1/2} \overset{a.s.}{=} \mathrm{o}(1). \tag{3.1}$$

If $|\mathrm{eigen}(\mathbf{B})| \leq 1$ *and* $j \geq 1$ *then*

$$\left\{ \sum_{t=1}^{T} (\mathbf{X}_t | D_t)^{\otimes 2} \right\}^{-1/2} \sum_{t=1}^{T} (\mathbf{X}_t | D_t) \varepsilon_{t+j}' \overset{a.s.}{=} \mathrm{O}(\log T). \tag{3.2}$$

It holds

$$\left(\sum_{t=1}^{T}\tilde{U}_t^{\otimes 2}\right)^{-1/2}\sum_{t=1}^{T}\tilde{U}_t\varepsilon_{t+j}' \overset{a.s.}{=} \mathrm{O}\left\{(\log\log T)^{1/2}\right\}, \tag{3.3}$$

$$\liminf_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^{\otimes 2} \overset{a.s.}{>} 0, \qquad \limsup_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^{\otimes 2} \overset{a.s.}{<} \infty, \tag{3.4}$$

$$\liminf_{T\to\infty} T^{-1}\log\lambda_{\min}\left(\sum_{t=1}^{T}W_t^{\otimes 2}\right) \overset{a.s.}{=} 2\log\min\left|\mathrm{eigen}\left(\mathbf{W}\right)\right|, \tag{3.5}$$

$$\liminf_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{X}_t|D_t\right)^{\otimes 2} \overset{a.s.}{>} 0. \tag{3.6}$$

$$\left\{\sum_{t=1}^{T}\left(\mathbf{X}_t|D_t\right)^{\otimes 2}\right\}^{-1/2}\sum_{t=1}^{T}\left(\mathbf{X}_t|D_t\right)\varepsilon_{t+j}' \overset{a.s.}{=} \mathrm{o}\left(T^{1/4}\right), \qquad \textit{for } j\geq 1. \tag{3.7}$$

For the results (3.1) $-$ (3.6) *it suffices that* $\gamma > 0$ *in Assumption* (2.1)*.*

Proof of Lemma 3.2. The results follow from Nielsen (2001). In particular, (3.1) follows from Example 8.4, (3.2) and (3.7) from Theorem 9.1, (3.3) from Theorem 7.7, (3.4) from Theorem 5.1, (3.5) from Corollary 6.2, and (3.6) from Corollary 8.5. ■

The Lemma 3.1 is now proved in a few steps. The first is an algebraic result showing that a k-th order partial correlation can be rewritten as the sum of two correlations and a first order partial correlation.

Lemma 3.3 *The statistic* $TS_{01}S_{11}^{-1}S_{10}$ *can be written as*

$$\begin{aligned}
&\sum_{t=1}^{T}\varepsilon_t\left(\mathbf{X}_{t-2}|D_t\right)'\left\{\sum_{t=1}^{T}\left(\mathbf{X}_{t-2}|D_t\right)^{\otimes 2}\right\}^{-1}\sum_{t=1}^{T}\left(\mathbf{X}_{t-2}|D_t\right)\varepsilon_t' \\
&-\sum_{t=1}^{T}\varepsilon_t\left(\mathbf{X}_{t-1}|D_t\right)'\left\{\sum_{t=1}^{T}\left(\mathbf{X}_{t-2}|D_t\right)^{\otimes 2}\right\}^{-1}\sum_{t=1}^{T}\left(\mathbf{X}_{t-2}|D_t\right)\varepsilon_t' \\
&+\sum_{t=1}^{T}\varepsilon_t\left(\varepsilon_{t-1}|\mathbf{X}_{t-2},D_t\right)'\left\{\sum_{t=1}^{T}\left(\varepsilon_{t-1}|\mathbf{X}_{t-2},D_t\right)^{\otimes 2}\right\}^{-1}\sum_{t=1}^{T}\left(\varepsilon_{t-1}|\mathbf{X}_{t-2},D_t\right)\varepsilon_t'.
\end{aligned}$$

Proof of Lemma 3.3. By the formula for partitioned inversion $TS_{01}S_{11}^{-1}S_{10}$ equals

$$\sum_{t=1}^{T}\varepsilon_t\left(X_{t-k}|\mathbf{X}_{t-1},D_t\right)'\left\{\sum_{t=1}^{T}\left(X_{t-k}|\mathbf{X}_{t-1},D_t\right)^{\otimes 2}\right\}^{-1}\sum_{t=1}^{T}\left(X_{t-k}|\mathbf{X}_{t-1},D_t\right)\varepsilon_t'$$

$$= \sum_{t=1}^{T} \varepsilon_t \left(\begin{array}{c} \mathbf{X}_{t-1} \\ X_{t-k} \end{array} \middle| D_t \right)' \left\{ \sum_{t=1}^{T} \left(\begin{array}{c} \mathbf{X}_{t-1} \\ X_{t-k} \end{array} \middle| D_t \right)^{\otimes 2} \right\}^{-1} \sum_{t=1}^{T} \left(\begin{array}{c} \mathbf{X}_{t-1} \\ X_{t-k} \end{array} \middle| D_t \right) \varepsilon_t'$$
$$- \sum_{t=1}^{T} \varepsilon_t \left(\mathbf{X}_{t-1} | D_t \right)' \left(\sum_{t=1}^{T} \left(\mathbf{X}_{t-1} | D_t \right)^{\otimes 2} \right)^{-1} \sum_{t=1}^{T} \left(\mathbf{X}_{t-1} | D_t \right) \varepsilon_t'.$$

Noting that $(\mathbf{X}_{t-1}', X_{t-k}')' = (X_{t-1}', \mathbf{X}_{t-2}')$ a repeated use of the formula for partitioned inversion leads to

$$\sum_{t=1}^{T} \varepsilon_t \left(\mathbf{X}_{t-2} | D_t \right)' \left\{ \sum_{t=1}^{T} \left(\mathbf{X}_{t-2} | D_t \right)^{\otimes 2} \right\}^{-1} \sum_{t=1}^{T} \left(\mathbf{X}_{t-2} | D_t \right) \varepsilon_t'$$
$$- \sum_{t=1}^{T} \varepsilon_t \left(\mathbf{X}_{t-1} | D_t \right)' \left\{ \sum_{t=1}^{T} \left(\mathbf{X}_{t-2} | D_t \right)^{\otimes 2} \right\}^{-1} \sum_{t=1}^{T} \left(\mathbf{X}_{t-2} | D_t \right) \varepsilon_t'$$
$$+ \sum_{t=1}^{T} \varepsilon_t \left(X_{t-1} | \mathbf{X}_{t-2}, D_t \right)' \left\{ \sum_{t=1}^{T} \left(X_{t-1} | \mathbf{X}_{t-2}, D_t \right)^{\otimes 2} \right\}^{-1} \sum_{t=1}^{T} \left(X_{t-1} | \mathbf{X}_{t-2}, D_t \right) \varepsilon_t'.$$

The desired result now follows since $(X_{t-1} | \mathbf{X}_{t-2}, D_t) = (\varepsilon_{t-1} | \mathbf{X}_{t-2}, D_t)$ by the model equation (1.1) and the property $D_t = \mathbf{D} D_{t-1}$. ■

The uncorrelatedness of the sample correlations mentioned in (3.1) in Lemma 3.2 implies that the first two terms in Lemma 3.3 can be rewritten as

$$\begin{aligned} & \sum_{t=1}^{T} \varepsilon_t \left(\mathbf{X}_{t-j} | D_t \right)' \left\{ \sum_{t=1}^{T} \left(\mathbf{X}_{t-j} | D_t \right)^{\otimes 2} \right\}^{-1} \sum_{t=1}^{T} \left(\mathbf{X}_{t-j} | D_t \right) \varepsilon_t' \\ \overset{a.s.}{=} & \sum_{t=1}^{T} \varepsilon_t \left(U_{t-j} | D_t \right)' \left\{ \sum_{t=1}^{T} \left(U_{t-j} | D_t \right)^{\otimes 2} \right\}^{-1} \sum_{t=1}^{T} \left(U_{t-j} | D_t \right) \varepsilon_t' \\ & + \sum_{t=1}^{T} \varepsilon_t \left(V_{t-j} | D_t \right)' \left\{ \sum_{t=1}^{T} \left(V_{t-j} | D_t \right)^{\otimes 2} \right\}^{-1} \sum_{t=1}^{T} \left(V_{t-j} | D_t \right) \varepsilon_t' \\ & + \sum_{t=1}^{T} \varepsilon_t \left(W_{t-j} | D_t \right)' \left\{ \sum_{t=1}^{T} \left(W_{t-j} | D_t \right)^{\otimes 2} \right\}^{-1} \sum_{t=1}^{T} \left(W_{t-j} | D_t \right) \varepsilon_t' + \mathrm{o}(1) \end{aligned} \tag{3.8}$$

The next step is to see that the components involving W cancel each other.

Lemma 3.4 *Suppose Assumptions* (2.1), (2.2), (2.3) *are satisfied. Then*

$$\sum_{t=1}^{T} \varepsilon_t \left(W_{t-2} | D_t \right)' \left\{ \sum_{t=1}^{T} \left(W_{t-2} | D_t \right)^{\otimes 2} \right\}^{-1} \sum_{t=1}^{T} \left(W_{t-2} | D_t \right) \varepsilon_t'$$
$$- \sum_{t=1}^{T} \varepsilon_t \left(W_{t-1} | D_t \right)' \left\{ \sum_{t=1}^{T} \left(W_{t-1} | D_t \right)^{\otimes 2} \right\}^{-1} \sum_{t=1}^{T} \left(W_{t-1} | D_t \right) \varepsilon_t' \overset{a.s.}{=} \mathrm{o}(1).$$

Proof of Lemma 3.4. It is first argued that

$$\sum_{t=1}^{T}(W_{t-1}|D_t)^{\otimes 2} = \sum_{t=1}^{T}(e_{W,t-1}+\mathbf{W}W_{t-2}|D_t)^{\otimes 2} \overset{a.s.}{=} \sum_{t=1}^{T}(\mathbf{W}W_{t-2}|D_t)^{\otimes 2}\left\{1+\mathrm{o}\left(T^{-1/2}\right)\right\}.$$

The model equation gives the first equality. Next, (3.5) together with (3.7) imply $\{\sum_{t=1}^{T}(\mathbf{W}W_{t-2}|D_t)^{\otimes 2}\}^{-1}\sum_{t=1}^{T}(\mathbf{W}W_{t-2}|D_t)\, e'_{W,t-1} = \mathrm{o}(T^{-1/2})$ *a.s.*, whereas (3.5) and (3.4) imply $\{\sum_{t=1}^{T}(\mathbf{W}W_{t-2}|D_t)^{\otimes 2}\}^{-1}\sum_{t=1}^{T}e^{\otimes 2}_{W,t-1} = \mathrm{o}_{\mathsf{P}}(T^{-1/2})$.

In the same way, using first the model equation and then (3.4) together with (3.2) for $\dim \mathbf{D} = 0$ show

$$\sum_{t=1}^{T}(W_{t-1}|D_t)\,\varepsilon'_t \overset{a.s.}{=} \sum_{t=1}^{T}(\mathbf{W}W_{t-2}|D_t)\,\varepsilon'_t + \mathrm{o}\left(T^{\eta+1/2}\right).$$

In combination, these two results together with (3.5), (3.7) give

$$\begin{aligned}
&\left\{\sum_{t=1}^{T}(W_{t-1}|D_t)^{\otimes 2}\right\}^{-1/2}\sum_{t=1}^{T}(W_{t-1}|D_t)\,\varepsilon'_t \\
\overset{a.s.}{=}\; &\left\{\sum_{t=1}^{T}(\mathbf{W}W_{t-2}|D_t)^{\otimes 2}\right\}^{-1/2}\left\{\sum_{t=1}^{T}(\mathbf{W}W_{t-2}|D_t)\,\varepsilon'_t + \mathrm{o}\left(T^{\eta+1/2}\right)\right\}\left\{1+\mathrm{o}\left(T^{-1/2}\right)\right\} \\
\overset{a.s.}{=}\; &\left\{\sum_{t=1}^{T}(W_{t-2}|D_t)^{\otimes 2}\right\}^{-1/2}\sum_{t=1}^{T}(W_{t-2}|D_t)\,\varepsilon'_t + \mathrm{o}\left(T^{-1/2}\right).
\end{aligned}$$

Squaring this expression and using (3.7) gives the desired result. ■

The proof of Lemma 3.1 can now be given.

Proof of Lemma 3.1. In combination the above Lemmas 3.3, 3.4 show that for asymptotic purposes $TS_{01}S_{11}^{-1}S_{10}$ equals

$$\begin{aligned}
&\sum_{t=1}^{T}\varepsilon_t\,(U_{t-2}|D_t)'\left\{\sum_{t=1}^{T}(U_{t-2}|D_t)^{\otimes 2}\right\}^{-1}\sum_{t=1}^{T}(U_{t-2}|D_t)\,\varepsilon'_t \\
&-\sum_{t=1}^{T}\varepsilon_t\,(U_{t-1}|D_t)'\left\{\sum_{t=1}^{T}(U_{t-1}|D_t)^{\otimes 2}\right\}^{-1}\sum_{t=1}^{T}(U_{t-1}|D_t)\,\varepsilon'_t \\
&+\sum_{t=1}^{T}\varepsilon_t\,(\varepsilon_{t-1}|\mathbf{X}_{t-2},D_t)'\left\{\sum_{t=1}^{T}(\varepsilon_{t-1}|\mathbf{X}_{t-2},D_t)^{\otimes 2}\right\}^{-1}\sum_{t=1}^{T}(\varepsilon_{t-1}|\mathbf{X}_{t-2},D_t)\,\varepsilon'_t + R_T + \mathrm{o}\,(1)\,.
\end{aligned}$$

In the first two terms $(U_{t-j}|D_t)$ can be replaced by $(\tilde{U}_t|D_t)$ because of (2.5). The regressions on D_t in the first two equations and on D_t, V_t and W_t in the last equation can be ignored for asymptotic purposes because of (3.7). The result then follows using Lemma 3.3. ■

A more compact version of the result in Lemma 3.1 can be proved under an additional assumption to the innovations.

Lemma 3.5 *Suppose Assumptions* (2.1), (2.2), (2.3) *are satisfied and* $T^{-1}\sum_{t=1}^{T}\varepsilon_t^{\otimes 2} \to \Omega$ *a.s. Then there exists an* $\{(p+\dim U)\times p\}$*-matrix* C *with full column rank so* $Y_t = C'(\varepsilon_t', U_{t-1}')'$ *and*

$$TS_{01}S_{11}^{-1}S_{10} \overset{a.s.}{=} \sum_{t=1}^{T}\varepsilon_t Y_{t-1}'\left(\sum_{t=1}^{T}Y_{t-1}^{\otimes 2}\right)^{-1}\sum_{t=1}^{T}Y_{t-1}\varepsilon_t' + R_T + \mathrm{o}\left(1\right).$$

Proof of Lemma 3.5. The additional condition $T^{-1}\sum_{t=1}^{T}\varepsilon_t^{\otimes 2} \to \Omega$ *a.s.* ensures that the exists a positive definite F so

$$\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}\varepsilon_t^{\otimes 2} & 0\\ 0 & \tilde{U}_{t-1}^{\otimes 2}\end{pmatrix} \overset{a.s.}{=} \begin{pmatrix}\Omega & 0\\ 0 & F\end{pmatrix}\left\{1+\mathrm{o}\left(1\right)\right\}, \tag{3.9}$$

see Nielsen (2001, Theorem 5.1). Recall that the innovation for $\tilde{U}_t$ is found by $M\left(\varepsilon_t', 0\right)' = (e_{U,t}', e_{V,t}', e_{W,t}')'$, where M is some real, full rank matrix. Introducing the notation M_{up} for the top left $(\dim U \times p)$-block of M it then holds

$$\tilde{U}_t = \mathbf{U}\tilde{U}_{t-1} + e_{U,t} = (M_{up}, \mathbf{U})\begin{pmatrix}\varepsilon_t\\ \tilde{U}_{t-1}\end{pmatrix},$$

where $(M_{up}, \mathbf{U})$ is $\{\dim U \times (p+\dim U)\}$ and has full row rank. By (3.7) it follows that

$$\begin{aligned} F &\overset{a.s.}{=} \frac{1}{T}\sum_{t=1}^{T}\tilde{U}_{t-1}^{\otimes 2} + \mathrm{o}\left(1\right) \overset{a.s.}{=} (M_{up}, \mathbf{U})\left\{\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}\varepsilon_t\\ \tilde{U}_{t-1}\end{pmatrix}^{\otimes 2}\right\}\begin{pmatrix}M_{up}'\\ \mathbf{U}'\end{pmatrix} + \mathrm{o}\left(1\right)\\ &\overset{a.s.}{=} (M_{up}, \mathbf{U})\left\{\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}\varepsilon_t^{\otimes 2} & 0\\ 0 & \tilde{U}_{t-1}^{\otimes 2}\end{pmatrix}\right\}\begin{pmatrix}M_{up}'\\ \mathbf{U}'\end{pmatrix} + \mathrm{o}\left(1\right). \end{aligned} \tag{3.10}$$

Now, choose a $\{(p+\dim U)\times p\}$-matrix C with full column rank, so $\{C, (M_{u1}, \mathbf{U})'\}$ is regular, and

$$\left\{\begin{matrix}(M_{up}, \mathbf{U})\\ C'\end{matrix}\right\}\begin{pmatrix}\Omega & 0\\ 0 & F\end{pmatrix}\left\{\begin{pmatrix}M_{up}'\\ \mathbf{U}'\end{pmatrix}, C\right\} = \begin{pmatrix}F & 0\\ 0 & G\end{pmatrix},$$

for some positive definite matrix G, defined as the limit of $T^{-1}\sum_{t=1}^{T} Y_t^{\otimes 2}$ where $Y_t = C'(\varepsilon_t', \tilde{U}_{t-1}')'$. Using the full rank linear transformation $\{C, (M_{up}, \mathbf{U})'\}$ and the above results (3.9), (3.10) it is then seen that

$$\begin{aligned}
&\sum_{t=1}^{T} \varepsilon_t \begin{pmatrix} \varepsilon_{t-1} \\ \tilde{U}_{t-2} \end{pmatrix}' \left\{ \sum_{t=1}^{T} \begin{pmatrix} \varepsilon_{t-1}^{\otimes 2} & 0 \\ 0 & \tilde{U}_{t-2}^{\otimes 2} \end{pmatrix} \right\}^{-1} \sum_{t=1}^{T} \begin{pmatrix} \varepsilon_{t-1} \\ \tilde{U}_{t-2} \end{pmatrix} \varepsilon_t' \\
&\overset{a.s.}{=} \sum_{t=1}^{T} \varepsilon_t \begin{pmatrix} U_{t-1} \\ Y_{t-1} \end{pmatrix}' \begin{pmatrix} F & 0 \\ 0 & G \end{pmatrix}^{-1} \sum_{t=1}^{T} \begin{pmatrix} U_{t-1} \\ Y_{t-1} \end{pmatrix} \varepsilon_t' + \mathrm{o}(1).
\end{aligned}$$

The desired result follows by subtracting $\sum_{t=1}^{T} \varepsilon_t U_{t-1}' \left(\sum_{t=1}^{T} U_{t-1}^{\otimes 2} \right)^{-1} \sum_{t=1}^{T} U_{t-1}\varepsilon_t'$. ■

The order of magnitude of the remainder term R_T is given in the following Lemma.

Lemma 3.6 *Suppose Assumptions* (2.1), (2.2), (2.3) *are satisfied.*
(i) *It holds* $R_T = \mathrm{O}(\log T)$ *a.s.*
(ii) *If Assumption* (2.8) *holds then* $R_T = \mathrm{o}(1)$ *a.s.*
(iii) *If Assumption* (2.6) *holds then* $R_T = \mathrm{o}_{\mathsf{P}}(1)$.

Proof of Lemma 3.6. (i) The property (3.2) in Lemma 3.2 immediately shows $R_T = \mathrm{O}(\log T)$ *a.s.*

(ii) as in the proof of Lemma 3.4 with (3.5), (3.7) replaced by (2.8), (3.2), respectively, it can be proved that

$$\begin{aligned}
\sum_{t=1}^{T} (V_{t-1}|D_t)^{\otimes 2} &\overset{a.s.}{=} \sum_{t=1}^{T} (\mathbf{V}V_{t-2}|D_t)^{\otimes 2} \left\{ 1 + \mathrm{o}\left(T^{-\eta}\right) \right\}, \\
\sum_{t=1}^{T} (V_{t-1}|D_t)\, \varepsilon_t' &\overset{a.s.}{=} \sum_{t=1}^{T} (\mathbf{V}V_{t-2}|D_t)\, \varepsilon_t' + \mathrm{o}\left(T^{\eta+1/2}\right).
\end{aligned}$$

The argument is then finished as in the proof of Lemma 3.4.

(iii) follows as (ii) using (2.7) instead of (2.8) ■

3.2 *Proofs of main results*

Theorem 2.1 is a consequence of Lemma 3.5 and a Central Limit Theorem.

Proof of Theorem 2.1. The condition $T^{-1}\sum_{t=1}^{T} \varepsilon_t^{\otimes 2} \to \Omega$ in Lemma 3.5 is satisfied because of Assumption (2.6), see Nielsen (2001, Theorem 9.5). By Lemma 3.6,*iii* it holds $R_T = \mathrm{o}_{\mathsf{P}}(1)$. The result then follows from Brown's (1971) Central Limit Theorem, see Nielsen (2001, Theorem 10.2). ■

The Theorems 2.2, 2.3, 2.5 follow from the final two Lemmas, of which the first gives lower bounds for $f(T)$ and the second gives an upper bound for $f(T)$.

Lemma 3.7 *Suppose Assumptions* (2.1), (2.2), (2.3) *are satisfied.*
(i) *If Assumption* (2.6) *holds and* $f(T) \to \infty$ *then* $\mathsf{P}(\hat{k}_0 > k_0) \to 0$.
(ii) *If* $f(T)/\log T \to \infty$ *then* $\limsup_{T\to\infty} \hat{k}_0 \leq k_0$ *a.s.*
(iii) *If Assumption* (2.8) *holds and* $f(T)/\log\log T \to \infty$ *then* $\limsup_{T\to\infty} \hat{k}_0 \leq k_0$ *a.s.*
(iv) *If Assumption* (2.8) *holds and* ε_t *is stationary and ergodic then* $\limsup_{T\to\infty} \hat{k}_0 \leq k_0$ *a.s. if and only if* $\limsup_{T\to\infty}(2\log\log T)^{-1} f(T) > 1$.

Proof of Lemma 3.7. Let $j \geq k_0$. It then holds that

$$\Phi_{j+1} - \Phi_j = \log\det\left(\hat{\Omega}_{j+1}\hat{\Omega}_j^{-1}\right) + T^{-1}f(T) = -T^{-1}LR(A_{j+1}) + T^{-1}f(T).$$

The order of magnitude of $LR(A_{j+1})$ therefore has to be evaluated.

(i) Theorem 2.1 with k replaced by $j+1$ shows that $LR(A_{j+1}) = \mathrm{O}_{\mathsf{P}}(1)$.

(ii) Lemmas 3.1, 3.6,i and (3.2) of Lemma 3.2 show $LR(A_{j+1}) = \mathrm{O}(\log T)$ *a.s.*

(iii) Lemma 3.6,ii implies that $R_T = \mathrm{o}(1)$ *a.s.* and (3.3) then shows $LR(A_{j+1}) = \mathrm{O}(\log\log T)$ *a.s.*

(iv) The Ergodic Theorem shows $T^{-1}\sum_{t=1}^T \varepsilon_t^{\otimes 2} \to \Omega$ *a.s.* and therefore Lemma 3.5 applies with $R_T = \mathrm{o}(1)$ because of Lemma 3.6,ii. The desired result then follows from the Law of Iteraterated Logarithms by Heyde and Scott (1973, Corollary 2) and Hannan (1980, p. 1076-1077). See Quinn (1980) for details. ■

Lemma 3.8 *Suppose Assumptions* (2.1), (2.2), (2.3) *are satisfied. If* $f(T) = \mathrm{o}(T)$ *then* $\liminf_{T\to\infty} \hat{k}_0 \geq k_0$ *a.s.*

Proof of Lemma 3.8. Let $j < k_0$. The condition $f(T) = \mathrm{o}(T)$ shows that

$$\Phi_j - \Phi_{k_0} = \log\det\left\{I + \left(\hat{\Omega}_j - \hat{\Omega}_{k_0}\right)\hat{\Omega}_{k_0}^{-1}\right\} + \mathrm{o}(1).$$

By (3.4) in Lemma 3.2 it holds that $\hat{\Omega}_{k_0}^{-1} = \mathrm{O}(1)$ *a.s.* Thus it suffices to argue that $\liminf \lambda_{\max}(\hat{\Omega}_j - \hat{\Omega}_{k_0}) > 0$. Defining $\mathbf{Y}_t = (X'_{t-j}, \ldots, X'_{t-k_0})'$ it holds

$$\hat{\Omega}_j - \hat{\Omega}_{k_0} = \frac{1}{T}\sum_{t=1}^T X_t (\mathbf{Y}_{t-1}|\mathbf{X}_{t-1}, D_t)' \left\{\sum_{t=1}^T (\mathbf{Y}_{t-1}|\mathbf{X}_{t-1}, D_t)^{\otimes 2}\right\}^{-1} \sum_{t=1}^T (\mathbf{Y}_{t-1}|\mathbf{X}_{t-1}, D_t) X'_t.$$

Define $\mathbf{A} = A_{j+1}, \ldots, A_{k_0}$ which is non-zero since $A_{k_0} \neq 0$ and note that

$$\begin{aligned}
&\frac{1}{T^{1/2}}\left\{\sum_{t=1}^T (\mathbf{Y}_{t-1}|\mathbf{X}_{t-1}, D_t)^{\otimes 2}\right\}^{-1/2} \sum_{t=1}^T (\mathbf{Y}_{t-1}|\mathbf{X}_{t-1}, D_t) X'_t \\
= \;& \frac{1}{T^{1/2}}\left\{\sum_{t=1}^T (\mathbf{Y}_{t-1}|\mathbf{X}_{t-1}, D_t)^{\otimes 2}\right\}^{-1/2} \sum_{t=1}^T (\mathbf{Y}_{t-1}|\mathbf{X}_{t-1}, D_t) \varepsilon'_t \\
& + \left\{\frac{1}{T}\sum_{t=1}^T (\mathbf{Y}_{t-1}|\mathbf{X}_{t-1}, D_t)^{\otimes 2}\right\}^{1/2} \mathbf{A}'.
\end{aligned}$$

The first of these terms is of order $\mathrm{o}(1)$ *a.s.* by (3.7), whereas the second term has a positive or infinite limes inferior because of (3.6), see also Nielsen (2001, Lemma 8.7).
■

References

Brown, B.M. (1971). Martingale Central Limit Theorems. *Annals of Mathematical Statistics* 42, 59-66.

Chan, N.H. and Wei, C.Z. (1988) Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics*, **16**, 367-401.

Donsker, M.D. and Varadhan, S.R.S. (1977) On Laws of Iterated Logarithms for local times. *Communications on Pure and Applied Mathematics*, **30**, 707-753.

Hannan, E.J. (1980) The estimation of the order of an ARMA process. *Annals of Statistics*, **8**, 1071-1081.

Hannan, E.J. and Quinn, B.G. (1979) The determination of the order of an autoregression. *Journal of the Royal Statistical Society* B, **41**, 190-195.

Herstein, I.N. (1975). *Topics in Algebra*, 2nd edition. New York: Wiley.

Heyde, C.C. and Scott, D.J. (1973) Invariance principles for the law of the iterated logarithm for martingales and processes with stationary increments. *Annals of Probability* 1, 428-436.

Lai, T.L. and Wei, C.Z. (1985) Asymptotic properties of multivariate weighted sums with applications to stochastic regression in linear dynamic systems. In P.R. Krishnaiah, ed., *Multivariate Analysis VI*, Elsevier Science Publishers, 375-393.

Lütkepohl, H. (1991) *Introduction to Multiple Time Series Analysis*. Berlin: Springer-Verlag.

Nielsen, B. (2001). Asymptotic properties of least squares statistics in general vector autoregressive models. See `http://www.nuff.ox.ac.uk/users/nielsen`.

Paulsen, J. (1984) Order determination of multivariate autoregressive time series with unit roots. *Journal of Time Series Analysis*, **5**, 115-127.

Quinn, B. G. (1980) Order determination for a multivariate autoregression. *Journal of the Royal Statistical Society* B, **42**, 182-185.

Schwarz, G. (1978) Estiamating the dimension of a model. *Annals of Statistics*, **6**, 461-464.

Tsay, R.S. (1984) Order selection in nonstationary autoregressive models. *Annals of Statistics*, **12**, 1425-1433.

Wei, C.Z. (1985) Asymptotic properties of least-squares estimates in stochastic regression models. *Annals of Statistics* 13, 1498-1508.