Recursive estimation and order determination of space-time autoregressive processes

Antunes, Ana Monica C. and Quinn, Barry, G and Subba Rao, T

2006

MIMS EPrint: **2006.37**

ISSN 1749-9097

Recursive Estimation and Order Determination of Space-Time Autoregressive Processes

Ana Monica C. Antunes, Barry G. Quinn & T. Subba Rao

First version: 23 September 2005

Research Report No. 4, 2005, Probability and Statistics Group
School of Mathematics, The University of Manchester

Recursive estimation and order determination of space-time autoregressive processes

By ANA MÓNICA C. ANTUNES*

Department of Mathematics

University of Manchester Institute of Science and Technology (UMIST)

PO Box 88, Manchester M60 1QD, UK

a.antunes@postgrad.umist.ac.uk

BARRY G. QUINN

Department of Statistics, Macquarie University

Sydney, NSW, 2109, Australia

bquinn@efs.mq.edu.au

And T. SUBBA RAO

Department of Mathematics

UMIST

tata.subbarao@umist.ac.uk

Abstract

Space-time autoregressive moving average models may be used for time series measured at the same times in a number of locations. In this paper we propose a recursive algorithm for estimating space-time autoregressive (AR) models. We also propose an information criterion for estimating the model order, and prove its strong consistency. The methods are illustrated using both simulated and real data. The real data corresponds to hourly carbon monoxide (CO) concentrations recorded in September 1995 at four different locations in Venice.

Key words: AIC, asymptotic normality, BIC, consistency, law of the iterated logarithm, order selection, recursive estimation, space-time AR process, Yule-Walker equations.

*Corresponding Author. Fax:+44 161 200 3669

1 Introduction

In many areas of research one is interested in the statistical analysis of measurements of one quantity taken at various locations (space) and times. Examples include air pollution concentrations at various locations in a city at different times, temperatures in a region, wind speeds and the spread of epidemics. Different methods have been proposed for the analysis of real data sets (see, for example, Mardia et al. (1998), Wickle & Cressie(1999), Guttorp et al. (1994), Shaddick & Wakefield (2002), Haslett & Raftery (1989)). (*Elaborate on how they have been used?*)

(*First STAR or STARMA model used when? Was it the same as in* (1)?) In this paper we shall consider space-time autoregressions (Pfeifer & Deutsch (1980)). Let $X(s_i, t)$ $(i = 1, ..., N,\ t = 1, ..., T)$ denote a random variable associated with location s_i and time t. Let $\mathbf{X}(t) \equiv (X(s_1; t), ..., X(s_N; t))'$, $t \in \mathbb{Z}$. $\{\mathbf{X}(t)\}$ is said to be a space-time AR process of order k if $\mathbf{X}(t)$ satisfies a difference equation of the form

$$\mathbf{X}(t) + \sum_{j=1}^{k} (\phi_j I_N + \psi_j W)\, \mathbf{X}(t-j) = \varepsilon(t) \tag{1}$$

where the scalars ϕ_j, ψ_j $(j = 1, ..., k)$ are unknown parameters. I_N is the N-dimensional identity matrix, W is a non-zero $N \times N$ known weighting matrix, with diagonal entries 0, without loss of generality) and off-diagonal entries related to the distances between the sites. Each row sum of W is normalised to 1, again without loss of generality. We assume that $\{\varepsilon(t)\}$ is weakly stationary, with

$$E\{\varepsilon(t)\} = 0 \tag{2}$$

and

$$E\left\{\varepsilon(t)\, \varepsilon(t+s)'\right\} = \begin{cases} \sigma^2 I_N & ; \quad s = 0 \\ 0 & ; \quad s \neq 0. \end{cases} \tag{3}$$

It is further assumed that the zeros of

$$\det\left\{ I_N + \sum_{j=1}^{k} (\phi_j I_N + \psi_j W)\, z^j \right\} \tag{4}$$

are outside the unit circle, which ensures that there exists a weakly stationary process which satisfies (1) .

Fields of application of the above models include meteorology (Subba Rao & Antunes(2004)), criminology (Pfeifer & Deutsch (1980)), ecology (Epperson (2000), Stoffer(1986)), transportation studies (Garrido (2000)) and many others. A space-time AR process is a very parsimonious multivariate autoregression. It is this parsimony, however, which prevents the use of algorithms such as the Whittle(1963) algorithm to estimate the unknown parameters. However, we can use the principal ideas of Whittle(1963) and Quinn (1980) (see also Hannan & Deistler (1988, sec.6, ch. 5)) to develop recursive equations for the estimation of these parameters. The Levinson-Durbin and Whittle algorithms have a natural place in the estimation of uinvariate and multivariate autoregressive order (references). In this paper we also introduce an information criterion to estimate the order k of the process and demonstrate its consistency.

In Section 2, we derive Yule-Walker relations for the space-time AR process. Using these equations we obtain recursive relations for estimating the parameters of the model. The sampling properties of the estimators are discussed in Section 3. The problem of estimating the order of the model is considered in Section 4. For clarity of explanation the proofs of the results obtained in Sections 2-4 are given in a separate Section 5). We illustrate the methodology with simulated data in Section 6. In Section 7 we use hourly carbon monoxide (CO) concentrations recorded in 1995 at four locations in Venice to illustrate our estimation procedures.

2 Yule-Walker equations and recursive algorithm

Here we obtain Yule-Walker like equations for space-time AR processes. These will be not the usually-specified ones, which represent a 1:1 relationship between autocovariances and system parameters, because of the fact that the autoregressive matrices are specified by the $(2k+1)$ parameters σ^2 and the ϕ_k and ψ_k. For $j \in \mathbb{Z}$, let

$$\gamma_j = E\left\{\mathbf{X}'(t)\,\mathbf{X}(t+j)\right\},$$
$$\pi_j = E\left\{\mathbf{X}'(t)\,W\mathbf{X}(t+j)\right\},$$
$$\lambda_j = E\left\{\mathbf{X}'(t)\,W'W\mathbf{X}(t+j)\right\}.$$

Put $\sigma_k^2 = \sigma^2$ and

$$\begin{aligned}
\Phi_k^{(k)} &= [\phi_1, ..., \phi_k]', \\
\Psi_k^{(k)} &= [\psi_1, ..., \psi_k]', \\
U_k &= \left\{\gamma_{i-j}\right\}_{i,j=1,...,k}, \\
V_k &= \left\{\pi_{i-j}\right\}_{i,j=1,...,k}, \\
F_k &= \left\{\lambda_{i-j}\right\}_{i,j=1,...,k}, \\
\Gamma_k &= \begin{bmatrix} \gamma_1 & \cdots & \gamma_k \end{bmatrix}', \\
\Lambda_k &= \begin{bmatrix} \lambda_1 & \cdots & \lambda_k \end{bmatrix}', \\
\Pi_k &= \begin{bmatrix} \pi_1 & \cdots & \pi_k \end{bmatrix}', \\
\Pi_{-k} &= \begin{bmatrix} \pi_{-1} & \cdots & \pi_{-k} \end{bmatrix}'.
\end{aligned}$$

THEOREM 1 *Let* $\{\mathbf{X}(t)\}$ *be a space-time AR process satisfying* (1) *where* $\{\varepsilon(t)\}$ *satisfies* (2) *and* (3). *Then*

$$\begin{bmatrix} U_k & V_k \\ V_k' & F_k \end{bmatrix} \begin{bmatrix} \Phi_k^{(k)} \\ \Psi_k^{(k)} \end{bmatrix} = - \begin{bmatrix} \Gamma_k \\ \Pi_{-k} \end{bmatrix} \tag{5}$$

$$N\sigma_k^2 = \gamma_0 + \Gamma_k'\Phi_k^{(k)} + \Pi_{-k}'\Psi_k^{(k)} \tag{6}$$

Proof: See Section 5.

That $\Phi_k^{(k)}$ and $\Psi_k^{(k)}$ can easily be calculated solving (5) follows from the following lemma.

LEMMA 2 *The $2k \times 2k$ matrix*

$$\Omega_k = \begin{bmatrix} U_k & V_k \\ V_k' & F_k \end{bmatrix}$$

is invertible for any k.

Proof: See Section 5.

For given k, the determination of $\Phi_k^{(k)}$ and $\Psi_k^{(k)}$ requires the inversion of the $2k$ by $2k$ matrix Ω_k. The computational load can be reduced by using the Töplitz structure of Ω_k, a fact which is utilised in the Levinson-Durbin and Whittle algorithms. Moreover, as is the case with these algorithms, at each recursive step, information about the system order is revealed. In the next theorem we introduce a recursive, computationally efficient algorithm for solving (5). It should be recalled that the Whittle recursion involves the invention of 'forward' autoregressive parameters. Our recursion will need these as well as another 'augmented' set of parameters.

Let $\Delta_k^{(k)}$, $\Theta_k^{(k)}$, $\underline{\Phi}_k^{(k)}$, $\underline{\Psi}_k^{(k)}$, $\underline{\Delta}_k^{(k)}$, $\underline{\Theta}_k^{(k)}$ satisfy

$$\begin{bmatrix} U_k & V_k \\ V_k' & F_k \end{bmatrix} \begin{bmatrix} \Phi_k^{(k)} & \Delta_k^{(k)} \\ \Psi_k^{(k)} & \Theta_k^{(k)} \end{bmatrix} = - \begin{bmatrix} \Gamma_k & \Pi_k \\ \Pi_{-k} & \Lambda_k \end{bmatrix}$$

and

$$\begin{bmatrix} U_k & V_k' \\ V_k & F_k \end{bmatrix} \begin{bmatrix} \underline{\Phi}_k^{(k)} & \underline{\Delta}_k^{(k)} \\ \underline{\Psi}_k^{(k)} & \underline{\Theta}_k^{(k)} \end{bmatrix} = - \begin{bmatrix} \Gamma_k & \Pi_{-k} \\ \Pi_k & \Lambda_k \end{bmatrix}. \tag{7}$$

For $k = 0, 1, \ldots,$ put

$$\begin{bmatrix} \Phi_{k+1}^{(k+1)} & \Delta_{k+1}^{(k+1)} \\ \Psi_{k+1}^{(k+1)} & \Theta_{k+1}^{(k+1)} \end{bmatrix} = \begin{bmatrix} \Phi_k^{(k+1)} & \Delta_k^{(k+1)} \\ \phi_{k+1}^{(k+1)} & \delta_{k+1}^{(k+1)} \\ \Psi_k^{(k+1)} & \Theta_k^{(k+1)} \\ \psi_{k+1}^{(k+1)} & \theta_{k+1}^{(k+1)} \end{bmatrix},$$

$$\begin{bmatrix} \underline{\Phi}_{k+1}^{(k+1)} & \underline{\Delta}_{k+1}^{(k+1)} \\ \underline{\Psi}_{k+1}^{(k+1)} & \underline{\Theta}_{k+1}^{(k+1)} \end{bmatrix} = \begin{bmatrix} \underline{\Phi}_k^{(k+1)} & \underline{\Delta}_k^{(k+1)} \\ \underline{\phi}_{k+1}^{(k+1)} & \underline{\delta}_{k+1}^{(k+1)} \\ \underline{\Psi}_k^{(k+1)} & \underline{\Theta}_k^{(k+1)} \\ \underline{\psi}_{k+1}^{(k+1)} & \underline{\theta}_{k+1}^{(k+1)} \end{bmatrix},$$

where $\Phi_k^{(k+1)}, \Psi_k^{(k+1)}, \Delta_k^{(k+1)}, \Theta_k^{(k+1)}, \underline{\Phi}_k^{(k+1)}, \underline{\Psi}_k^{(k+1)}, \underline{\Delta}_k^{(k+1)}$ and $\underline{\Theta}_k^{(k+1)}$ are $k \times 1$ and $\phi_{k+1}^{(k+1)}, \psi_{k+1}^{(k+1)}, \delta_{k+1}^{(k+1)}, \theta_{k+1}^{(k+1)}, \underline{\phi}_{k+1}^{(k+1)}, \underline{\psi}_{k+1}^{(k+1)}, \underline{\delta}_{k+1}^{(k+1)}$ and $\underline{\theta}_{k+1}^{(k+1)}$ are scalar. For any vector $x = \begin{bmatrix} x_1 & \cdots & x_m \end{bmatrix}'$, let $\widetilde{x} = \begin{bmatrix} x_m & \cdots & x_1 \end{bmatrix}'$, i.e. x with its elements reversed. Let

$$S_0 = T_0 = \begin{bmatrix} \gamma_0 & \pi_0 \\ \pi_0 & \lambda_0 \end{bmatrix}, P_0 = Q_0' = \begin{bmatrix} \gamma_1 & \pi_1 \\ \pi_{-1} & \lambda_1 \end{bmatrix}$$

and for $k \geq 1$, let

$$P_k = \begin{bmatrix} \gamma_{k+1} & \pi_{k+1} \\ \pi_{-k-1} & \lambda_{k+1} \end{bmatrix} + \begin{bmatrix} \widetilde{\Gamma}_k' & \widetilde{\Pi}_k' \\ \widetilde{\Pi}_{-k}' & \widetilde{\Lambda}_k' \end{bmatrix} \begin{bmatrix} \Phi_k^{(k)} & \Delta_k^{(k)} \\ \Psi_k^{(k)} & \Theta_k^{(k)} \end{bmatrix}, \tag{8}$$

$$Q_k = \begin{bmatrix} \gamma_{k+1} & \pi_{-k-1} \\ \pi_{k+1} & \lambda_{k+1} \end{bmatrix} + \begin{bmatrix} \widetilde{\Gamma}_k' & \widetilde{\Pi}_{-k}' \\ \widetilde{\Pi}_k' & \widetilde{\Lambda}_k' \end{bmatrix} \begin{bmatrix} \underline{\Phi}_k^{(k)} & \underline{\Delta}_k^{(k)} \\ \underline{\Psi}_k^{(k)} & \underline{\Theta}_k^{(k)} \end{bmatrix}, \tag{9}$$

$$S_k = \begin{bmatrix} \gamma_0 & \pi_0 \\ \pi_0 & \lambda_0 \end{bmatrix} + \begin{bmatrix} \Gamma_k' & \Pi_{-k}' \\ \Pi_k' & \Lambda_k' \end{bmatrix} \begin{bmatrix} \Phi_k^{(k)} & \Delta_k^{(k)} \\ \Psi_k^{(k)} & \Theta_k^{(k)} \end{bmatrix}, \tag{10}$$

$$T_k = \begin{bmatrix} \gamma_0 & \pi_0 \\ \pi_0 & \lambda_0 \end{bmatrix} + \begin{bmatrix} \Gamma_k' & \Pi_k' \\ \Pi_{-k}' & \Lambda_k' \end{bmatrix} \begin{bmatrix} \underline{\Phi}_k^{(k)} & \underline{\Delta}_k^{(k)} \\ \underline{\Psi}_k^{(k)} & \underline{\Theta}_k^{(k)} \end{bmatrix} \tag{11}$$

THEOREM 3 *For $k \geq 0$,*

$$\begin{bmatrix} \phi_{k+1}^{(k+1)} & \delta_{k+1}^{(k+1)} \\ \psi_{k+1}^{(k+1)} & \theta_{k+1}^{(k+1)} \end{bmatrix} = -T_k^{-1} P_k, \begin{bmatrix} \underline{\phi}_{k+1}^{(k+1)} & \underline{\delta}_{k+1}^{(k+1)} \\ \underline{\psi}_{k+1}^{(k+1)} & \underline{\theta}_{k+1}^{(k+1)} \end{bmatrix} = -S_k^{-1} Q_k$$

and for $k \geq 1$,

$$\begin{bmatrix} \Phi_k^{(k+1)} & \Delta_k^{(k+1)} \\ \Psi_k^{(k+1)} & \Theta_k^{(k+1)} \end{bmatrix} = \begin{bmatrix} \Phi_k^{(k)} & \Delta_k^{(k)} \\ \Psi_k^{(k)} & \Theta_k^{(k)} \end{bmatrix} + \begin{bmatrix} \widetilde{\underline{\Phi}}_k^{(k)} & \widetilde{\underline{\Delta}}_k^{(k)} \\ \widetilde{\underline{\Psi}}_k^{(k)} & \widetilde{\underline{\Theta}}_k^{(k)} \end{bmatrix} \begin{bmatrix} \phi_{k+1}^{(k+1)} & \delta_{k+1}^{(k+1)} \\ \psi_{k+1}^{(k+1)} & \theta_{k+1}^{(k+1)} \end{bmatrix},$$

$$\begin{bmatrix} \underline{\Phi}_k^{(k+1)} & \underline{\Delta}_k^{(k+1)} \\ \underline{\Psi}_k^{(k+1)} & \underline{\Theta}_k^{(k+1)} \end{bmatrix} = \begin{bmatrix} \underline{\Phi}_k^{(k)} & \underline{\Delta}_k^{(k)} \\ \underline{\Psi}_k^{(k)} & \underline{\Theta}_k^{(k)} \end{bmatrix} + \begin{bmatrix} \widetilde{\Phi}_k^{(k)} & \widetilde{\Delta}_k^{(k)} \\ \widetilde{\Psi}_k^{(k)} & \widetilde{\Theta}_k^{(k)} \end{bmatrix} \begin{bmatrix} \underline{\phi}_{k+1}^{(k+1)} & \underline{\delta}_{k+1}^{(k+1)} \\ \underline{\psi}_{k+1}^{(k+1)} & \underline{\theta}_{k+1}^{(k+1)} \end{bmatrix}.$$

Proof: See Section 5.

The following theorem shows that S_k and T_k are covariance matrices and are invertible.

THEOREM 4 *Let $\varepsilon_k(t), u_k(t), \underline{\varepsilon}_k(t)$ and $\underline{u}_k(t)$ be the differences between $\mathbf{X}(t)$ and* *its predictors* $\sum_{j=1}^{k}\left(a_jI+b_jW\right)\mathbf{X}\left(t-j\right)$ *which respectively minimise*

$$E\left\{\mathbf{X}\left(t\right)-\sum_{j=1}^{k}\left(a_jI+b_jW\right)\mathbf{X}\left(t-j\right)\right\}'\left\{\mathbf{X}\left(t\right)-\sum_{j=1}^{k}\left(a_jI+b_jW\right)\mathbf{X}\left(t-j\right)\right\},$$

$$E\left\{W\mathbf{X}\left(t\right)-\sum_{j=1}^{k}\left(c_jI+d_jW\right)\mathbf{X}\left(t-j\right)\right\}'\left\{W\mathbf{X}\left(t\right)-\sum_{j=1}^{k}\left(c_jI+d_jW\right)\mathbf{X}\left(t-j\right)\right\},$$

$$E\left\{\mathbf{X}\left(t\right)-\sum_{j=1}^{k}\left(a_jI+b_jW\right)\mathbf{X}\left(t+j\right)\right\}'\left\{\mathbf{X}\left(t\right)-\sum_{j=1}^{k}\left(a_jI+b_jW\right)\mathbf{X}\left(t+j\right)\right\}$$

and

$$E\left\{W\mathbf{X}\left(t\right)-\sum_{j=1}^{k}\left(c_jI+d_jW\right)\mathbf{X}\left(t+j\right)\right\}'\left\{W\mathbf{X}\left(t\right)-\sum_{j=1}^{k}\left(c_jI+d_jW\right)\mathbf{X}\left(t+j\right)\right\}.$$

Then

$$S_k=E\left\{\begin{bmatrix}\varepsilon_k'\left(t\right)\\ u_k'\left(t\right)\end{bmatrix}\begin{bmatrix}\varepsilon_k\left(t\right) & u_k\left(t\right)\end{bmatrix}\right\},$$

$$T_k=E\left\{\begin{bmatrix}\underline{\varepsilon}_k'\left(t\right)\\ \underline{u}_k'\left(t\right)\end{bmatrix}\begin{bmatrix}\underline{\varepsilon}_k\left(t\right) & \underline{u}_k\left(t\right)\end{bmatrix}\right\}$$

and S_k *and* T_k *are of full rank for all* k*. Furthermore,* $N\sigma_k^2$ *is the* $(1,1)$ *entry of* S_k*.*

Proof: See Section 5.

The following lemma presents recursive expressions for S_k, T_k and σ_k^2. The equations (12) and (13) should be used instead of (10) and (11) when estimating, as the former equations are expected to be less prone to rounding problems. Similar relations exist for the Levinson-Durbin and Whittle algorithms.

LEMMA 5 *Let* S_k *and* T_k *be as defined above. Then*

$$S_{k+1}=S_k\left\{I-\begin{bmatrix}\underline{\phi}_{k+1}^{(k+1)} & \underline{\delta}_{k+1}^{(k+1)}\\ \underline{\psi}_{k+1}^{(k+1)} & \underline{\theta}_{k+1}^{(k+1)}\end{bmatrix}\begin{bmatrix}\phi_{k+1}^{(k+1)} & \delta_{k+1}^{(k+1)}\\ \psi_{k+1}^{(k+1)} & \theta_{k+1}^{(k+1)}\end{bmatrix}\right\}, \tag{12}$$

$$T_{k+1}=T_k\left\{I-\begin{bmatrix}\phi_{k+1}^{(k+1)} & \delta_{k+1}^{(k+1)}\\ \psi_{k+1}^{(k+1)} & \theta_{k+1}^{(k+1)}\end{bmatrix}\begin{bmatrix}\underline{\phi}_{k+1}^{(k+1)} & \underline{\delta}_{k+1}^{(k+1)}\\ \underline{\psi}_{k+1}^{(k+1)} & \underline{\theta}_{k+1}^{(k+1)}\end{bmatrix}\right\}. \tag{13}$$

Proof: See Section 5.

The following lemma reduces the computations needed in solving the recursive equations in Theorem 3.

LEMMA 6 *Let P_k and Q_k be as defined in (8) and (9), respectively. Then*

$$Q_k = P_k'.$$

Proof: See Section 5.

3 Parameter estimation for space-time AR processes

The above results describe exact relationships between the theoretical covariance and autocovariance matrices, and the true parameters. The above algorithm becomes an estimation algorithm when the theoretical covariances are replaced by consistent estimators. Let $\{\mathbf{X}(t)\}$ be a space-time process. Given $\{\mathbf{X}(t); 1 \leq t \leq T\}$, estimate γ_j, π_j and λ_j by the obvious moment estimators

$$\hat{\gamma}_j = T^{-1} \sum_{t=1}^{T-j} \mathbf{X}'(t)\,\mathbf{X}(t+j),$$
$$\hat{\pi}_j = T^{-1} \sum_{t=1}^{T-j} \mathbf{X}'(t)\,W\mathbf{X}(t+j)$$

and

$$\hat{\lambda}_j = T^{-1} \sum_{t=1}^{T-j} \mathbf{X}'(t)\,W'W\mathbf{X}(t+j),$$

respectively. Without loss of generality, we shall assume that the $\mathbf{X}(t)$ have mean zero, for, in practice, we shall always mean-correct them. This mean-correction will have no asymptotic effect, so we omit any further reference for notational simplicity.

The following additional weak assumptions will be made in order to develop the asymptotic properties of the estimators:

1. $\{\varepsilon(t)\}$ is strictly stationary and ergodic;

2. $E\{\varepsilon(t)|\,\mathcal{F}_{t-1}\} = 0$;

3. $E\left\{\varepsilon\left(t\right)\varepsilon'\left(t\right)\middle|\mathcal{F}_{t-1}\right\} = \sigma^2 I_N$;

4. $E\left\{\varepsilon_j^4\left(t\right)\right\} < \infty,\ j = 1, \ldots, N,$

where $\mathcal{F}_t$ is the σ-field generated by $\left\{\varepsilon\left(s\right); s \leq t\right\}$. In the following we denote generically by $\widehat{\theta}$ the estimator of θ obtained from the recursion by substituting $\hat{\gamma}_j$ for γ_j, etc.

THEOREM 7 *Let* $\left\{\mathbf{X}\left(t\right)\right\}$ *satisfy* (1) *and the above assumptions. Then*

1. $\widehat{\Phi}_k^{(k)}$ *and* $\widehat{\Psi}_k^{(k)}$ *converge almost surely to* $\Phi_k^{(k)}$ *and* $\Psi_k^{(k)}$;

2. *The distribution of* $\sqrt{T}\left[\ \left\{\widehat{\Phi}_k^{(k)} - \Phi_k^{(k)}\right\}'\ \ \left\{\widehat{\Psi}_k^{(k)} - \Psi_k^{(k)}\right\}'\ \right]'$ *converges, as* $T \to \infty$*, to the normal with mean zero and covariance matrix*

$$\sigma^2 \begin{bmatrix} U_k & V_k \\ V_k' & F_k \end{bmatrix}^{-1}.$$

Proof: See Section 5.

4 Estimating k

The above presupposes the true order k to be known. In practice, however, we shall need to estimate k. The following theorem describes the behaviour of a class of AIC-type estimators of k and establishes conditions under which these estimators are strongly and weakly consistent.

THEOREM 8 *Let* $\widehat{k}$ *be the minimiser of*

$$\phi_f\left(k\right) = NT\log\widehat{\sigma}_k^2 + 2kf\left(T\right)$$

over $k = 0, 1, \ldots, K$*, where* K *is known* a priori *to be larger than* k_0*, the true order, and* $f\left(T\right)/T \to 0$ *as* $T \to \infty$*. Then*

1. $\widehat{k}$ *converges to* k_0 *in probability if* $f\left(T\right)$ *diverges to* ∞;

2. If $f(T) = c$, for all T, then

$$\lim_{T\to\infty} \Pr\left\{\widehat{k} = k_0 + j\right\} = \begin{cases} 0 & ; \quad j < 0 \\ p_j q_{K-k_0-j} & ; \quad 0 \le j \le K - k_0, \end{cases}$$

where the p_j and q_j are generated by

$$\sum_{j=0}^{\infty} p_j z^j = \exp\left\{\sum_{j=1}^{\infty} j^{-1} z^j \Pr\left(\chi^2_{2j} > 2cj\right)\right\}$$

and

$$\sum_{j=0}^{\infty} q_j z^j = \exp\left\{\sum_{j=1}^{\infty} j^{-1} z^j \Pr\left(\chi^2_{2j} \le 2cj\right)\right\}.$$

In particular,

$$\begin{aligned}\lim_{T\to\infty} \Pr\left\{\widehat{k} = k_0\right\} &= q_{K-k_0} \\ &\to \exp\left\{-\sum_{j=1}^{\infty} j^{-1} \Pr\left(\chi^2_{2j} > 2cj\right)\right\},\end{aligned}$$

as $K \to \infty$.

3. $\widehat{k}$ converges to k_0 almost surely if

$$\liminf_{T} \frac{f(T)}{\log\log T} > 1,$$

and only if

$$\liminf_{T} \frac{f(T)}{\log\log T} \ge 1.$$

Thus, it can be seen that, when $f(T)$ is constant, the asymptotic probability of overestimating the order is greater than zero and the probability of underestimation is zero. Also, the probability of estimating the correct order is strictly less than 1. Hence the AIC estimator (which corresponds to the case where $f(T) = 2$) is not a consistent estimator of the order. The case

$$\liminf_{T} \frac{f(T)}{\log\log T} = 1$$

is of some interest but will not be dealt with here, especially as the law of the iterated logarithm used to prove these results (see Lemma 10 in the appendix) can only be expected to hold for extremely large values of T.

5 Proofs of theorems

Here we prove the results contained in the theorems and lemmas stated in Sections 2 - 4.

PROOF of Theorem 1

We motivate the normal equations and recursions by developing the Gaussian maximum likelihood procedure, i.e. the maximum likelihood estimators when $\{\varepsilon(t)\}$ is assumed to be Gaussian. We do not, however, need to assume that $\{\varepsilon(t)\}$ is Gaussian – we merely use the Gaussian likelihood to obtain the correct "normal equations", which will hold whether or not $\{\varepsilon(t)\}$ is Gaussian. The Gaussian maximum likelihood estimators of the parameters ϕ_j, ψ_j $(j = 1, ..., k)$ and σ^2 for a space-time AR process of order k can be obtained by maximizing the Gaussian log-likelihood function

$$l_k = -\frac{TN}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\varepsilon'(t)\,\varepsilon(t)$$

where $\varepsilon(t)$ is now understood to be the function of $\phi_1, ..., \phi_k,\ \psi_1, ..., \psi_k,\ \sigma^2$ and $\mathbf{X}(1), \ldots, \mathbf{X}(T)$ given by

$$\varepsilon(t) = \mathbf{X}(t) + \sum_{j=1}^{k}(\phi_j I + \psi_j W)\,\mathbf{X}(t-j)\,.$$

This is equiavlent to minimizing the sum of squares

$$\begin{aligned} S_k(\Phi, \Psi) &= \sum_{t=1}^{T}\varepsilon'(t)\,\varepsilon(t) \\ &= \sum_{t=1}^{T}\left\{\mathbf{X}(t) + \sum_{j=1}^{k}(\phi_j I + \psi_j W)\,\mathbf{X}(t-j)\right\}' \\ &\times\left\{\mathbf{X}(t) + \sum_{j=1}^{k}(\phi_j I + \psi_j W)\,\mathbf{X}(t-j)\right\} \end{aligned}$$

where $\Phi = [\phi_1, ..., \phi_k]'$ and $\Psi = [\psi_1, ..., \psi_k]'$, and then setting $TN\sigma^2$ equal to the minimum value of $S_k(\Phi, \Psi)$.

Since $S_k(\Phi, \Psi)$ is quadratic in Φ and Ψ, we can minimise $S_k(\Phi, \Psi)$ by differentiating with respect to ϕ_m, ψ_m $(m = 1, ..., k)$ and equating the derivatives to zero. Now,

$$\frac{\partial S_k(\Phi, \Psi)}{\partial \phi_m} = 2\sum_{t=1}^{T} \mathbf{X}'(t-m)\left\{\mathbf{X}(t) + \sum_{j=1}^{k}(\phi_j I + \psi_j W)\mathbf{X}(t-j)\right\}$$
$$\frac{\partial S_k(\Phi, \Psi)}{\partial \psi_m} = 2\sum_{t=1}^{T} \mathbf{X}'(t-m)\,W'\left\{\mathbf{X}(t) + \sum_{j=1}^{k}(\phi_j I + \psi_j W)\mathbf{X}(t-j)\right\}.$$

"Least squares" estimators may be obtained by solving these equations, which comprise a set of $2k$ linear equations in $2k$ unknowns. We wish, however, to develop a recursive algorithm such as the Levinson-Durbin and Whittle algorithms, and so need to obtain the correct formulation for the Yule-Walker relations for space time autoregressions. These equations, which are theoretical "normal equations", are obtained by equating the expectations of the right hand sides of the above to 0. We thus have for each $m = 1, \ldots, k$

$$\gamma_m + \sum_{j=1}^{k}(\phi_j\gamma_{m-j} + \psi_j\pi_{m-j}) = 0$$
$$\pi_{-m} + \sum_{j=1}^{k}(\phi_j\pi_{j-m} + \psi_j\lambda_{m-j}) = 0.$$

Stacking these equations from $m = 1$ to k, we obtain (5). Note that $\gamma_{-m} = \gamma_m$ and $\lambda_{-m} = \lambda_m$ but that $\pi_{-m} \neq \pi_m$. It is straightforward to verify that these equations hold true for any weakly stationary process $\{\mathbf{X}(t)\}$ satisfying (1). Moreover, from the above,

$$\begin{aligned}
N\sigma^2 &= E\left\{T^{-1}\sum_{t=1}^{T}\varepsilon'(t)\,\varepsilon(t)\right\} \\
&= \gamma_0 + \sum_{j=1}^{k}(\phi_j\gamma_j + \psi_j\pi_{-j}) + \sum_{j=1}^{k}\phi_j\left(\gamma_j + \sum_{i=1}^{k}\phi_i\gamma_{j-i} + \sum_{i=1}^{k}\psi_i\pi_{j-i}\right) \\
&+ \sum_{j=1}^{k}\psi_j\left(\pi_{-j} + \sum_{i=1}^{k}\phi_i\pi_{i-j} + \sum_{i=1}^{k}\psi_i\lambda_{j-i}\right) \\
&= \gamma_0 + \sum_{j=1}^{k}(\phi_j\gamma_j + \psi_j\pi_{-j})\,.
\end{aligned}$$

This completes the proof of the theorem.

PROOF of Lemma 2

Let α and β be arbitrary constant $k\times1$ vectors, with $\left[\begin{array}{cc}\alpha' & \beta'\end{array}\right] \neq 0$. The quadratic form

$$\left[\begin{array}{cc}\alpha' & \beta'\end{array}\right]\Omega_k\left[\begin{array}{c}\alpha \\ \beta\end{array}\right]$$

is the expectation of

$$\begin{aligned}
&\sum_{i,j=1}^{k}\{\alpha_i\mathbf{X}'(t-i) + \beta_i\mathbf{X}'(t-i)\,W'\}\{\alpha_j\mathbf{X}(t-j) + \beta_jW\mathbf{X}(t-j)\} \\
&= Z'Z
\end{aligned}$$

where

$$Z = \sum_{j=1}^{k}Z_j$$

and

$$Z_j = \alpha_j\mathbf{X}(t-j) + \beta_jW\mathbf{X}(t-j)\,.$$

Thus Ω_k is non-negative definite and it is invertible unless $Z = 0$ almost everywhere, which occurs only when there is a linear relation amongst $\mathbf{X}(t-1),\ldots,\mathbf{X}(t-k)$ and therefore amongst the $\varepsilon(t)$. However, we have assumed that $\{\mathbf{X}(t)\}$ satisfies (1) and is weakly stationary, and that $\{\varepsilon(t)\}$ is an uncorrelated process. There can therefore be no linear relations and so Ω_k is invertible and positive definite.

PROOF of Theorem 3

From (5) , the normal equations of order $k+1$ may be written, using the notation in Theorem 3, as

$$\begin{bmatrix} U_k & \widetilde{\Gamma}_k & V_k & \widetilde{\Pi}_{-k} \\ \widetilde{\Gamma}'_k & \gamma_0 & \widetilde{\Pi}'_k & \pi_0 \\ V'_k & \widetilde{\Pi}_k & F_k & \widetilde{\Lambda}_k \\ \widetilde{\Pi}'_{-k} & \pi_0 & \widetilde{\Lambda}'_k & \lambda_0 \end{bmatrix} \begin{bmatrix} \Phi_k^{(k+1)} \\ \phi_{k+1}^{(k+1)} \\ \Psi_k^{(k+1)} \\ \psi_{k+1}^{(k+1)} \end{bmatrix} = - \begin{bmatrix} \Gamma_k \\ \gamma_{k+1} \\ \Pi_{-k} \\ \pi_{-k-1} \end{bmatrix}.$$

We now proceed to derive the recursions for calculating $\Phi_{k+1}^{(k+1)}$ and $\Psi_{k+1}^{(k+1)}$ in terms of $\Phi_k^{(k)}$ and $\Psi_k^{(k)}$. The above equations may be written as the two equations

$$\begin{bmatrix} U_k & V_k \\ V'_k & F_k \end{bmatrix} \begin{bmatrix} \Phi_k^{(k+1)} \\ \Psi_k^{(k+1)} \end{bmatrix} + \begin{bmatrix} \widetilde{\Gamma}_k & \widetilde{\Pi}_{-k} \\ \widetilde{\Pi}_k & \widetilde{\Lambda}_k \end{bmatrix} \begin{bmatrix} \phi_{k+1}^{(k+1)} \\ \psi_{k+1}^{(k+1)} \end{bmatrix} = - \begin{bmatrix} \Gamma_k \\ \Pi_{-k} \end{bmatrix} \tag{14}$$

and

$$\begin{bmatrix} \widetilde{\Gamma}'_k & \widetilde{\Pi}'_k \\ \widetilde{\Pi}'_{-k} & \widetilde{\Lambda}'_k \end{bmatrix} \begin{bmatrix} \Phi_k^{(k+1)} \\ \Psi_k^{(k+1)} \end{bmatrix} + \begin{bmatrix} \gamma_0 & \pi_0 \\ \pi_0 & \lambda_0 \end{bmatrix} \begin{bmatrix} \phi_{k+1}^{(k+1)} \\ \psi_{k+1}^{(k+1)} \end{bmatrix} = - \begin{bmatrix} \gamma_{k+1} \\ \pi_{-k-1} \end{bmatrix}. \tag{15}$$

Equations (14) and (5) give

$$\begin{aligned} &\begin{bmatrix} \Phi_k^{(k+1)} \\ \Psi_k^{(k+1)} \end{bmatrix} \\ &= - \begin{bmatrix} U_k & V_k \\ V'_k & F_k \end{bmatrix}^{-1} \begin{bmatrix} \Gamma_k \\ \Pi_{-k} \end{bmatrix} - \begin{bmatrix} U_k & V_k \\ V'_k & F_k \end{bmatrix}^{-1} \begin{bmatrix} \widetilde{\Gamma}_k & \widetilde{\Pi}_{-k} \\ \widetilde{\Pi}_k & \widetilde{\Lambda}_k \end{bmatrix} \begin{bmatrix} \phi_{k+1}^{(k+1)} \\ \psi_{k+1}^{(k+1)} \end{bmatrix} \\ &= \begin{bmatrix} \Phi_k^{(k)} \\ \Psi_k^{(k)} \end{bmatrix} - \begin{bmatrix} U_k & V_k \\ V'_k & F_k \end{bmatrix}^{-1} \begin{bmatrix} \widetilde{\Gamma}_k & \widetilde{\Pi}_{-k} \\ \widetilde{\Pi}_k & \widetilde{\Lambda}_k \end{bmatrix} \begin{bmatrix} \phi_{k+1}^{(k+1)} \\ \psi_{k+1}^{(k+1)} \end{bmatrix}. \end{aligned}$$

Thus, from (15) , we have

$$\begin{aligned} &\left(\begin{bmatrix} \gamma_0 & \pi_0 \\ \pi_0 & \lambda_0 \end{bmatrix} - \begin{bmatrix} \widetilde{\Gamma}'_k & \widetilde{\Pi}'_k \\ \widetilde{\Pi}'_{-k} & \widetilde{\Lambda}'_k \end{bmatrix} \begin{bmatrix} U_k & V_k \\ V'_k & F_k \end{bmatrix}^{-1} \begin{bmatrix} \widetilde{\Gamma}_k & \widetilde{\Pi}_{-k} \\ \widetilde{\Pi}_k & \widetilde{\Lambda}_k \end{bmatrix} \right) \begin{bmatrix} \phi_{k+1}^{(k+1)} \\ \psi_{k+1}^{(k+1)} \end{bmatrix} \\ &= - \left(\begin{bmatrix} \gamma_{k+1} \\ \pi_{-k-1} \end{bmatrix} + \begin{bmatrix} \widetilde{\Gamma}'_k & \widetilde{\Pi}'_k \\ \widetilde{\Pi}'_{-k} & \widetilde{\Lambda}'_k \end{bmatrix} \begin{bmatrix} \Phi_k^{(k)} \\ \Psi_k^{(k)} \end{bmatrix} \right), \end{aligned}$$

or

$$
\begin{aligned}
&\begin{bmatrix} \phi_{k+1}^{(k+1)} \\ \psi_{k+1}^{(k+1)} \end{bmatrix} \\
&= -\left(\begin{bmatrix} \gamma_0 & \pi_0 \\ \pi_0 & \lambda_0 \end{bmatrix} - \begin{bmatrix} \widetilde{\Gamma}'_k & \widetilde{\Pi}'_k \\ \widetilde{\Pi}'_{-k} & \widetilde{\Lambda}'_k \end{bmatrix} \begin{bmatrix} U_k & V_k \\ V'_k & F_k \end{bmatrix}^{-1} \begin{bmatrix} \widetilde{\Gamma}_k & \widetilde{\Pi}_{-k} \\ \widetilde{\Pi}_k & \widetilde{\Lambda}_k \end{bmatrix} \right)^{-1} P_{k1} \\
&= -\left(\begin{bmatrix} \gamma_0 & \pi_0 \\ \pi_0 & \lambda_0 \end{bmatrix} - \begin{bmatrix} \Gamma'_k & \Pi'_k \\ \Pi'_{-k} & \Lambda'_k \end{bmatrix} \begin{bmatrix} U_k & V'_k \\ V_k & F_k \end{bmatrix}^{-1} \begin{bmatrix} \Gamma_k & \Pi_{-k} \\ \Pi_k & \Lambda_k \end{bmatrix} \right)^{-1} P_{k1} \qquad (16)
\end{aligned}
$$

where P_{k1} is the first column of P_k. The above simplification results using the following reasoning: Suppose

$$
\begin{bmatrix} U_k & V_k \\ V'_k & F_k \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} \widetilde{\Gamma}_k & \widetilde{\Pi}_{-k} \\ \widetilde{\Pi}_k & \widetilde{\Lambda}_k \end{bmatrix}, \qquad (17)
$$

where a, b, c and d are of the same dimension. We can obtain

$$
\begin{bmatrix} \Gamma_k & \Pi_{-k} \\ \Pi_k & \Lambda_k \end{bmatrix}
$$

on the right hand side of (17) by reversing the rows of

$$
\begin{bmatrix} U_k & V_k \end{bmatrix}
$$

and

$$
\begin{bmatrix} V'_k & F_k \end{bmatrix}.
$$

To obtain recognisable matrices, we then need to reverse the relevant columns. U_k and F_k remain unchanged, after reversal of both rows and columns, as they are Töplitz, while V_k is changed to V'_k. We preserve (17) by reversing a, b, c and d. Thus (17) is the same as

$$
\begin{bmatrix} U_k & V'_k \\ V_k & F_k \end{bmatrix} \begin{bmatrix} \widetilde{a} & \widetilde{b} \\ \widetilde{c} & \widetilde{d} \end{bmatrix} = \begin{bmatrix} \Gamma_k & \Pi_{-k} \\ \Pi_k & \Lambda_k \end{bmatrix},
$$

and we therefore have

$$\begin{bmatrix} \widetilde{\Gamma}'_k & \widetilde{\Pi}'_k \\ \widetilde{\Pi}'_{-k} & \widetilde{\Lambda}'_k \end{bmatrix} \begin{bmatrix} U_k & V_k \\ V'_k & F_k \end{bmatrix}^{-1} \begin{bmatrix} \widetilde{\Gamma}_k & \widetilde{\Pi}_{-k} \\ \widetilde{\Pi}_k & \widetilde{\Lambda}_k \end{bmatrix}$$
$$= \begin{bmatrix} \widetilde{\Gamma}'_k & \widetilde{\Pi}'_k \\ \widetilde{\Pi}'_{-k} & \widetilde{\Lambda}'_k \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$
$$= \begin{bmatrix} \Gamma'_k & \Pi'_k \\ \Pi'_{-k} & \Lambda'_k \end{bmatrix} \begin{bmatrix} \widetilde{a} & \widetilde{b} \\ \widetilde{c} & \widetilde{d} \end{bmatrix}$$
$$= \begin{bmatrix} \Gamma'_k & \Pi'_k \\ \Pi'_{-k} & \Lambda'_k \end{bmatrix} \begin{bmatrix} U_k & V'_k \\ V_k & F_k \end{bmatrix}^{-1} \begin{bmatrix} \Gamma_k & \Pi_{-k} \\ \Pi_k & \Lambda_k \end{bmatrix}.$$

Noting the differences between

$$-\begin{bmatrix} U_k & V'_k \\ V_k & F_k \end{bmatrix}^{-1} \begin{bmatrix} \Gamma_k & \Pi_{-k} \\ \Pi_k & \Lambda_k \end{bmatrix}$$

and the solution to (5) :

$$\begin{bmatrix} U_k & V_k \\ V'_k & F_k \end{bmatrix}^{-1} \begin{bmatrix} \Gamma_k \\ \Pi_{-k} \end{bmatrix} = -\begin{bmatrix} \Phi_k^{(k)} \\ \Psi_k^{(k)} \end{bmatrix},$$

it is clear that we need a double recursion and that we also need to augment the equations. Using the notation defined in the statement of the theorem, we augment (5) to the equation

$$\begin{bmatrix} U_k & V_k \\ V'_k & F_k \end{bmatrix} \begin{bmatrix} \Phi_k^{(k)} & \Delta_k^{(k)} \\ \Psi_k^{(k)} & \Theta_k^{(k)} \end{bmatrix} = -\begin{bmatrix} \Gamma_k & \Pi_k \\ \Pi_{-k} & \Lambda_k \end{bmatrix} \tag{18}$$

and in the light of (16) , define the 'forward' equations

$$\begin{bmatrix} U_k & V'_k \\ V_k & F_k \end{bmatrix} \begin{bmatrix} \underline{\Phi}_k^{(k)} & \underline{\Delta}_k^{(k)} \\ \underline{\Psi}_k^{(k)} & \underline{\Theta}_k^{(k)} \end{bmatrix} = -\begin{bmatrix} \Gamma_k & \Pi_{-k} \\ \Pi_k & \Lambda_k \end{bmatrix}. \tag{19}$$

Using similar methods as used to obtain (14) and (15) , we obtain

$$\begin{bmatrix} \Phi_k^{(k+1)} & \Delta_k^{(k+1)} \\ \Psi_k^{(k+1)} & \Theta_k^{(k+1)} \end{bmatrix}$$

$$= \begin{bmatrix} \Phi_k^{(k)} & \Delta_k^{(k)} \\ \Psi_k^{(k)} & \Theta_k^{(k)} \end{bmatrix} - \begin{bmatrix} U_k & V_k \\ V_k' & F_k \end{bmatrix}^{-1} \begin{bmatrix} \widetilde{\Gamma}_k & \widetilde{\Pi}_{-k} \\ \widetilde{\Pi}_k & \widetilde{\Lambda}_k \end{bmatrix} \begin{bmatrix} \phi_{k+1}^{(k+1)} & \delta_{k+1}^{(k+1)} \\ \psi_{k+1}^{(k+1)} & \theta_{k+1}^{(k+1)} \end{bmatrix} \tag{20}$$

$$= \begin{bmatrix} \Phi_k^{(k)} & \Delta_k^{(k)} \\ \Psi_k^{(k)} & \Theta_k^{(k)} \end{bmatrix} + \begin{bmatrix} \underline{\widetilde{\Phi}}_k^{(k)} & \underline{\widetilde{\Delta}}_k^{(k)} \\ \underline{\widetilde{\Psi}}_k^{(k)} & \underline{\widetilde{\Theta}}_k^{(k)} \end{bmatrix} \begin{bmatrix} \phi_{k+1}^{(k+1)} & \delta_{k+1}^{(k+1)} \\ \psi_{k+1}^{(k+1)} & \theta_{k+1}^{(k+1)} \end{bmatrix}, \tag{21}$$

$$\begin{bmatrix} \underline{\Phi}_k^{(k+1)} & \underline{\Delta}_k^{(k+1)} \\ \underline{\Psi}_k^{(k+1)} & \underline{\Theta}_k^{(k+1)} \end{bmatrix}$$

$$= \begin{bmatrix} \underline{\Phi}_k^{(k)} & \underline{\Delta}_k^{(k)} \\ \underline{\Psi}_k^{(k)} & \underline{\Theta}_k^{(k)} \end{bmatrix} - \begin{bmatrix} U_k & V_k' \\ V_k & F_k \end{bmatrix}^{-1} \begin{bmatrix} \widetilde{\Gamma}_k & \widetilde{\Pi}_k \\ \widetilde{\Pi}_{-k} & \widetilde{\Lambda}_k \end{bmatrix} \begin{bmatrix} \underline{\phi}_{k+1}^{(k+1)} & \underline{\delta}_{k+1}^{(k+1)} \\ \underline{\psi}_{k+1}^{(k+1)} & \underline{\theta}_{k+1}^{(k+1)} \end{bmatrix} \tag{22}$$

$$= \begin{bmatrix} \underline{\Phi}_k^{(k)} & \underline{\Delta}_k^{(k)} \\ \underline{\Psi}_k^{(k)} & \underline{\Theta}_k^{(k)} \end{bmatrix} + \begin{bmatrix} \widetilde{\Phi}_k^{(k)} & \widetilde{\Delta}_k^{(k)} \\ \widetilde{\Psi}_k^{(k)} & \widetilde{\Theta}_k^{(k)} \end{bmatrix} \begin{bmatrix} \phi_{k+1}^{(k+1)} & \delta_{k+1}^{(k+1)} \\ \psi_{k+1}^{(k+1)} & \theta_{k+1}^{(k+1)} \end{bmatrix},$$

$$\begin{bmatrix} \phi_{k+1}^{(k+1)} & \delta_{k+1}^{(k+1)} \\ \psi_{k+1}^{(k+1)} & \theta_{k+1}^{(k+1)} \end{bmatrix}$$

$$= -\left(\begin{bmatrix} \gamma_0 & \pi_0 \\ \pi_0 & \lambda_0 \end{bmatrix} - \begin{bmatrix} \Gamma_k' & \Pi_k' \\ \Pi_{-k}' & \Lambda_k' \end{bmatrix} \begin{bmatrix} U_k & V_k' \\ V_k & F_k \end{bmatrix}^{-1} \begin{bmatrix} \Gamma_k & \Pi_{-k} \\ \Pi_k & \Lambda_k \end{bmatrix} \right)^{-1} P_k$$

$$= -\left(\begin{bmatrix} \gamma_0 & \pi_0 \\ \pi_0 & \lambda_0 \end{bmatrix} + \begin{bmatrix} \Gamma_k' & \Pi_k' \\ \Pi_{-k}' & \Lambda_k' \end{bmatrix} \begin{bmatrix} \underline{\Phi}_k^{(k)} & \underline{\Delta}_k^{(k)} \\ \underline{\Psi}_k^{(k)} & \underline{\Theta}_k^{(k)} \end{bmatrix} \right)^{-1} P_k$$

$$= -T_k^{-1} P_k,$$

and

$$\left(\begin{bmatrix} \gamma_0 & \pi_0 \\ \pi_0 & \lambda_0 \end{bmatrix} - \begin{bmatrix} \widetilde{\Gamma}_k' & \widetilde{\Pi}_{-k}' \\ \widetilde{\Pi}_k' & \widetilde{\Lambda}_k' \end{bmatrix} \begin{bmatrix} U_k & V_k' \\ V_k & F_k \end{bmatrix}^{-1} \begin{bmatrix} \widetilde{\Gamma}_k & \widetilde{\Pi}_k \\ \widetilde{\Pi}_{-k} & \widetilde{\Lambda}_k \end{bmatrix} \right) \begin{bmatrix} \underline{\phi}_{k+1}^{(k+1)} & \underline{\delta}_{k+1}^{(k+1)} \\ \underline{\psi}_{k+1}^{(k+1)} & \underline{\theta}_{k+1}^{(k+1)} \end{bmatrix}$$

$$= -\left(\begin{bmatrix} \gamma_{k+1} & \pi_{-k-1} \\ \pi_{k+1} & \lambda_{k+1} \end{bmatrix} + \begin{bmatrix} \widetilde{\Gamma}_k' & \widetilde{\Pi}_{-k}' \\ \widetilde{\Pi}_k' & \widetilde{\Lambda}_k' \end{bmatrix} \begin{bmatrix} \underline{\Phi}_k^{(k)} & \underline{\Delta}_k^{(k)} \\ \underline{\Psi}_k^{(k)} & \underline{\Theta}_k^{(k)} \end{bmatrix} \right),$$

i.e.

$$\begin{bmatrix} \underline{\phi}_{k+1}^{(k+1)} & \underline{\delta}_{k+1}^{(k+1)} \\ \underline{\psi}_{k+1}^{(k+1)} & \underline{\theta}_{k+1}^{(k+1)} \end{bmatrix}$$

$$= -\left(\begin{bmatrix} \gamma_0 & \pi_0 \\ \pi_0 & \lambda_0 \end{bmatrix} - \begin{bmatrix} \Gamma'_k & \Pi'_{-k} \\ \Pi'_k & \Lambda'_k \end{bmatrix} \begin{bmatrix} U_k & V_k \\ V'_k & F_k \end{bmatrix}^{-1} \begin{bmatrix} \Gamma_k & \Pi_k \\ \Pi_{-k} & \Lambda_k \end{bmatrix} \right)^{-1} Q_k$$

$$= -\left(\begin{bmatrix} \gamma_0 & \pi_0 \\ \pi_0 & \lambda_0 \end{bmatrix} + \begin{bmatrix} \Gamma'_k & \Pi'_{-k} \\ \Pi'_k & \Lambda'_k \end{bmatrix} \begin{bmatrix} \Phi_k^{(k)} & \Delta_k^{(k)} \\ \Psi_k^{(k)} & \Theta_k^{(k)} \end{bmatrix} \right)^{-1} Q_k$$

$$= -S_k^{-1} Q_k.$$

The above equations are valid for $k \geq 1$. The recursion commences with the matrices

$$\begin{bmatrix} \phi_1^{(1)} & \delta_1^{(1)} \\ \psi_1^{(1)} & \theta_1^{(1)} \end{bmatrix}$$

and

$$\begin{bmatrix} \underline{\phi}_1^{(1)} & \underline{\delta}_1^{(1)} \\ \underline{\psi}_1^{(1)} & \underline{\theta}_1^{(1)} \end{bmatrix}.$$

From above, these are, respectively, given by

$$-\begin{bmatrix} U_1 & V_1 \\ V_1 & F_1 \end{bmatrix}^{-1} \begin{bmatrix} \Gamma_1 & \Pi_1 \\ \Pi_{-1} & \Lambda_1 \end{bmatrix} = -\begin{bmatrix} \gamma_0 & \pi_0 \\ \pi_0 & \lambda_0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma_1 & \pi_1 \\ \pi_{-1} & \lambda_1 \end{bmatrix}$$
$$= -T_0^{-1} P_0$$

and

$$-\begin{bmatrix} U_1 & V_1 \\ V_1 & F_1 \end{bmatrix}^{-1} \begin{bmatrix} \Gamma_1 & \Pi_{-1} \\ \Pi_1 & \Lambda_1 \end{bmatrix} = -\begin{bmatrix} \gamma_0 & \pi_0 \\ \pi_0 & \lambda_0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma_1 & \pi_{-1} \\ \pi_1 & \lambda_1 \end{bmatrix}$$
$$= -S_0^{-1} Q_0$$

This completes the proof.

PROOF of Theorem 4

Let $\varepsilon_k(t)$ be $\mathbf{X}(t) - \sum_{j=1}^{k}(a_j I + b_j W)\mathbf{X}(t-j)$, where the a_j and b_j are chosen to minimise

$$E\left\{\mathbf{X}(t) - \sum_{j=1}^{k}(a_j I + b_j W)\mathbf{X}(t-j)\right\}'\left\{\mathbf{X}(t) - \sum_{j=1}^{k}(a_j I + b_j W)\mathbf{X}(t-j)\right\}.$$

Then, since all relevant covariances exist, by differentiating and equating to zero we obtain

$$\begin{aligned}
&0\\
&= -\frac{1}{2}\frac{\partial}{\partial a_i}E\left\{\mathbf{X}(t) - \sum_{j=1}^{k}(a_j I + b_j W)\mathbf{X}(t-j)\right\}'\left\{\mathbf{X}(t) - \sum_{j=1}^{k}(a_j I + b_j W)\mathbf{X}(t-j)\right\}\\
&= E\left[\mathbf{X}'(t-i)\left\{\mathbf{X}(t) - \sum_{j=1}^{k}(a_j I + b_j W)\mathbf{X}(t-j)\right\}\right]\\
&= \gamma_i - \sum_{j=1}^{k}(a_j\gamma_{i-j} + b_j\pi_{i-j})
\end{aligned}$$

and

$$\begin{aligned}
&0\\
&= -\frac{1}{2}\frac{\partial}{\partial b_i}E\left\{\mathbf{X}(t) - \sum_{j=1}^{k}(a_j I + b_j W)\mathbf{X}(t-j)\right\}'\left\{\mathbf{X}(t) - \sum_{j=1}^{k}(a_j I + b_j W)\mathbf{X}(t-j)\right\}\\
&= E\left[\mathbf{X}'(t-i)W'\left\{\mathbf{X}(t) - \sum_{j=1}^{k}(a_j I + b_j W)\mathbf{X}(t-j)\right\}\right]\\
&= \pi_{-i} - \sum_{j=1}^{k}(a_j\pi_{j-i} + b_j\lambda_{j-i}).
\end{aligned}$$

Similarly, we obtain

$$\begin{aligned}
0 &= E\left[\mathbf{X}'(t-i)\left\{W\mathbf{X}(t) - \sum_{j=1}^{k}(c_j I + d_j W)\mathbf{X}(t-j)\right\}\right]\\
&= \pi_i - \sum_{j=1}^{k}(c_j\gamma_{i-j} + d_j\pi_{i-j})
\end{aligned}$$

and

$$
\begin{aligned}
0 &= E\left[\mathbf{X}'(t-i)\,W'\left\{W\mathbf{X}(t) - \sum_{j=1}^{k}(c_jI + d_jW)\,\mathbf{X}(t-j)\right\}\right] \\
&= \lambda_{-i} - \sum_{j=1}^{k}(c_j\pi_{j-i} + d_j\lambda_{j-i})\,.
\end{aligned}
$$

Thus, from (18), if $a = \begin{bmatrix} a_1 & \cdots & a_k \end{bmatrix}'$, $b = \begin{bmatrix} b_1 & \cdots & b_k \end{bmatrix}'$, $c = \begin{bmatrix} c_1 & \cdots & c_k \end{bmatrix}'$ and $d = \begin{bmatrix} d_1 & \cdots & d_k \end{bmatrix}'$, then

$$
\begin{bmatrix} a & c \\ b & d \end{bmatrix} = -\begin{bmatrix} \Phi_k^{(k)} & \Delta_k^{(k)} \\ \Psi_k^{(k)} & \Theta_k^{(k)} \end{bmatrix}.
$$

Hence, letting $\phi_j^{(k)}$ and $\psi_j^{(k)}$ be the jth components of $\Phi_k^{(k)}$ and $\Psi_k^{(k)}$, we obtain, using (5),

$$
\begin{aligned}
&E\left\{\varepsilon_k'(t)\,\varepsilon_k(t)\right\} \\
&= E\left\{\mathbf{X}(t) + \sum_{j=1}^{k}\left(\phi_j^{(k)}I + \psi_j^{(k)}W\right)\mathbf{X}(t-j)\right\}'\left\{\mathbf{X}(t) + \sum_{j=1}^{k}\left(\phi_j^{(k)}I + \psi_j^{(k)}W\right)\mathbf{X}(t-j)\right\} \\
&= \gamma_0 + \sum_{j=1}^{k}\left(\phi_j^{(k)}\gamma_j + \psi_j^{(k)}\pi_{-j}\right) + \sum_{j=1}^{k}\phi_j^{(k)}\left\{\gamma_j + \sum_{i=1}^{k}\left(\phi_i^{(k)}\gamma_{j-i} + \psi_j^{(k)}\pi_{i-j}\right)\right\} \\
&+ \sum_{j=1}^{k}\psi_j^{(k)}\left\{\pi_j + \sum_{i=1}^{k}\left(\phi_i^{(k)}\pi_{j-i} + \psi_j^{(k)}\lambda_{j-i}\right)\right\} \\
&= \gamma_0 + \Gamma_k'\Phi_k^{(k)} + \Pi_{-k}'\Psi_k^{(k)}.
\end{aligned}
$$

The above expression is just the $(1,1)$ element of S_k which is also equal to $N\sigma_k^2$. The same reasoning shows that $E\{\varepsilon_k'(t)\,u_k(t)\}$ and $E\{u_k'(t)\,u_k(t)\}$ are the $(1,2)$ and $(2,2)$ elements of S_k. The result for T_k follows using an identical argument, but using 'prediction' using the future, rather than the past.

We now show that S_k and T_k are (strictly) positive definite for each k. Suppose S_k has a zero eigenvalue. Then there exist scalars α and β, not both 0, such that

$$
\alpha\varepsilon_k(t) + \beta u_k(t) = 0,
$$

a.e. But, letting k_0 be the true order, we have

$$\alpha\varepsilon_k(t)+\beta u_k(t) = \alpha\left\{\mathbf{X}(t)-\sum_{j=1}^{k}(a_jI+b_jW)\mathbf{X}(t-j)\right\}$$
$$+\beta\left\{W\mathbf{X}(t)-\sum_{j=1}^{k}(c_jI+d_jW)\mathbf{X}(t-j)\right\}$$
$$=(\alpha I+\beta W)\mathbf{X}(t)-\sum_{j=1}^{k}\{(\alpha a_j+\beta c_j)I+(\alpha b_j+\beta d_j)W\}\mathbf{X}(t-j)$$
$$=(\alpha I+\beta W)\varepsilon(t)-(\alpha I+\beta W)\sum_{j=1}^{k_0}(\phi_jI+\psi_jW)\mathbf{X}(t-j)$$
$$-\sum_{j=1}^{k}\{(\alpha a_j+\beta c_j)I+(\alpha b_j+\beta d_j)W\}\mathbf{X}(t-j).$$

However, the last two terms (the sums) are linear in $\mathbf{X}(t-1),\ldots$ and therefore uncorrelated with $\varepsilon(t)$. Thus $\alpha\varepsilon_k(t)+\beta u_k(t)$ is zero, a.e. only if $(\alpha I+\beta W)\varepsilon(t)$ is zero. This in turn implies that

$$\Sigma = E\begin{bmatrix}\varepsilon'(t)\\ \varepsilon'(t)W'\end{bmatrix}\begin{bmatrix}\varepsilon(t) & W\varepsilon(t)\end{bmatrix}$$

is not of full rank. But

$$\Sigma=\sigma^2\begin{bmatrix}N & \operatorname{tr}W\\ \operatorname{tr}W & \operatorname{tr}(W'W)\end{bmatrix}$$
$$=\sigma^2\begin{bmatrix}N & 0\\ 0 & \operatorname{tr}(W'W)\end{bmatrix}.$$

Thus the S_k (and in turn the T_k, since $\det T_k=\det S_k$) are invertible as long as

$$\operatorname{tr}(W'W)\neq 0.$$

But

$$\operatorname{tr}(W'W)=\sum_{i,j=1}^{N}W_{ij}^2.$$

Thus

$$\operatorname{tr}(W'W) > 0,$$

with equality if and only if the W_{ij} are zero for all i and j. But we have ruled out this case, for otherwise $\{\mathbf{X}(t)\}$ would be a vector of uncorrelated autoregressions.

PROOF of Lemma 5

Although (10) and (11) define S_k and T_k, these formulae can sometimes cause problems in practice. Just as the Levinson-Durbin (Durbin (1960)) algorithm and Whittle(1963) recursions have alternative, more numerically stable formulae for the innovation variances and covariance matrices (see Quinn (1980), for example), there are alternative recursive formulae for the S_k and T_k. From (10) and (21) we have

$$\begin{aligned}
&S_{k+1} \\
&= \begin{bmatrix} \gamma_0 & \pi_0 \\ \pi_0 & \lambda_0 \end{bmatrix} + \begin{bmatrix} \Gamma'_{k+1} & \Pi'_{-k-1} \\ \Pi'_{k+1} & \Lambda'_{k+1} \end{bmatrix} \begin{bmatrix} \Phi^{(k+1)}_{k+1} & \Delta^{(k+1)}_{k+1} \\ \Psi^{(k+1)}_{k+1} & \Theta^{(k+1)}_{k+1} \end{bmatrix} \\
&= \begin{bmatrix} \gamma_0 & \pi_0 \\ \pi_0 & \lambda_0 \end{bmatrix} + \begin{bmatrix} \Gamma'_{k} & \Pi'_{-k} \\ \Pi'_{k} & \Lambda'_{k} \end{bmatrix} \begin{bmatrix} \Phi^{(k+1)}_{k} & \Delta^{(k+1)}_{k} \\ \Psi^{(k+1)}_{k} & \Theta^{(k+1)}_{k} \end{bmatrix} \\
&+ \begin{bmatrix} \gamma_{k+1} & \pi_{-k-1} \\ \pi_{k+1} & \lambda_{k+1} \end{bmatrix} \begin{bmatrix} \phi^{(k+1)}_{k+1} & \delta^{(k+1)}_{k+1} \\ \psi^{(k+1)}_{k+1} & \theta^{(k+1)}_{k+1} \end{bmatrix} \\
&= \begin{bmatrix} \gamma_0 & \pi_0 \\ \pi_0 & \lambda_0 \end{bmatrix} + \begin{bmatrix} \Gamma'_{k} & \Pi'_{-k} \\ \Pi'_{k} & \Lambda'_{k} \end{bmatrix} \begin{bmatrix} \Phi^{(k)}_{k} & \Delta^{(k)}_{k} \\ \Psi^{(k)}_{k} & \Theta^{(k)}_{k} \end{bmatrix} \\
&+ \begin{bmatrix} \Gamma'_{k} & \Pi'_{-k} \\ \Pi'_{k} & \Lambda'_{k} \end{bmatrix} \begin{bmatrix} \underline{\widetilde{\Phi}}^{(k)}_{k} & \underline{\widetilde{\Delta}}^{(k)}_{k} \\ \underline{\widetilde{\Psi}}^{(k)}_{k} & \underline{\widetilde{\Theta}}^{(k)}_{k} \end{bmatrix} \begin{bmatrix} \phi^{(k+1)}_{k+1} & \delta^{(k+1)}_{k+1} \\ \psi^{(k+1)}_{k+1} & \theta^{(k+1)}_{k+1} \end{bmatrix} \\
&+ \begin{bmatrix} \gamma_{k+1} & \pi_{-k-1} \\ \pi_{k+1} & \lambda_{k+1} \end{bmatrix} \begin{bmatrix} \phi^{(k+1)}_{k+1} & \delta^{(k+1)}_{k+1} \\ \psi^{(k+1)}_{k+1} & \theta^{(k+1)}_{k+1} \end{bmatrix}
\end{aligned}$$

$$= S_k + Q_k \begin{bmatrix} \phi_{k+1}^{(k+1)} & \delta_{k+1}^{(k+1)} \\ \psi_{k+1}^{(k+1)} & \theta_{k+1}^{(k+1)} \end{bmatrix}$$

$$= S_k - S_k \begin{bmatrix} \underline{\phi}_{k+1}^{(k+1)} & \underline{\delta}_{k+1}^{(k+1)} \\ \underline{\psi}_{k+1}^{(k+1)} & \underline{\theta}_{k+1}^{(k+1)} \end{bmatrix} \begin{bmatrix} \phi_{k+1}^{(k+1)} & \delta_{k+1}^{(k+1)} \\ \psi_{k+1}^{(k+1)} & \theta_{k+1}^{(k+1)} \end{bmatrix}$$

$$= S_k \left(I - \begin{bmatrix} \underline{\phi}_{k+1}^{(k+1)} & \underline{\delta}_{k+1}^{(k+1)} \\ \underline{\psi}_{k+1}^{(k+1)} & \underline{\theta}_{k+1}^{(k+1)} \end{bmatrix} \begin{bmatrix} \phi_{k+1}^{(k+1)} & \delta_{k+1}^{(k+1)} \\ \psi_{k+1}^{(k+1)} & \theta_{k+1}^{(k+1)} \end{bmatrix} \right).$$

Similarly,

$$T_{k+1} = T_k \left(I - \begin{bmatrix} \phi_{k+1}^{(k+1)} & \delta_{k+1}^{(k+1)} \\ \psi_{k+1}^{(k+1)} & \theta_{k+1}^{(k+1)} \end{bmatrix} \begin{bmatrix} \underline{\phi}_{k+1}^{(k+1)} & \underline{\delta}_{k+1}^{(k+1)} \\ \underline{\psi}_{k+1}^{(k+1)} & \underline{\theta}_{k+1}^{(k+1)} \end{bmatrix} \right).$$

PROOF of Lemma 6

By definition,

$$P_k = \begin{bmatrix} \gamma_{k+1} & \pi_{k+1} \\ \pi_{-k-1} & \lambda_{k+1} \end{bmatrix} + \begin{bmatrix} \widetilde{\Gamma}'_k & \widetilde{\Pi}'_k \\ \widetilde{\Pi}'_{-k} & \widetilde{\Lambda}'_k \end{bmatrix} \begin{bmatrix} \Phi_k^{(k)} & \Delta_k^{(k)} \\ \Psi_k^{(k)} & \Theta_k^{(k)} \end{bmatrix}$$

$$= \begin{bmatrix} \gamma_{k+1} & \pi_{k+1} \\ \pi_{-k-1} & \lambda_{k+1} \end{bmatrix} - \begin{bmatrix} \widetilde{\Gamma}'_k & \widetilde{\Pi}'_k \\ \widetilde{\Pi}'_{-k} & \widetilde{\Lambda}'_k \end{bmatrix} \begin{bmatrix} U_k & V_k \\ V'_k & F_k \end{bmatrix}^{-1} \begin{bmatrix} \Gamma_k & \Pi_k \\ \Pi_{-k} & \Lambda_k \end{bmatrix}$$

$$= \begin{bmatrix} \gamma_{k+1} & \pi_{k+1} \\ \pi_{-k-1} & \lambda_{k+1} \end{bmatrix} - \begin{bmatrix} \Gamma'_k & \Pi'_k \\ \Pi'_{-k} & \Lambda'_k \end{bmatrix} \begin{bmatrix} U_k & V'_k \\ V_k & F_k \end{bmatrix}^{-1} \begin{bmatrix} \widetilde{\Gamma}_k & \widetilde{\Pi}_k \\ \widetilde{\Pi}_{-k} & \widetilde{\Lambda}_k \end{bmatrix}$$

$$= \begin{bmatrix} \gamma_{k+1} & \pi_{k+1} \\ \pi_{-k-1} & \lambda_{k+1} \end{bmatrix} + \begin{bmatrix} \underline{\Phi}_k^{(k)} & \underline{\Delta}_k^{(k)} \\ \underline{\Psi}_k^{(k)} & \underline{\Theta}_k^{(k)} \end{bmatrix}' \begin{bmatrix} \widetilde{\Gamma}_k & \widetilde{\Pi}_k \\ \widetilde{\Pi}_{-k} & \widetilde{\Lambda}_k \end{bmatrix}$$

$$= \left(\begin{bmatrix} \gamma_{k+1} & \pi_{-k-1} \\ \pi_{k+1} & \lambda_{k+1} \end{bmatrix} + \begin{bmatrix} \widetilde{\Gamma}'_k & \widetilde{\Pi}'_{-k} \\ \widetilde{\Pi}'_k & \widetilde{\Lambda}'_k \end{bmatrix} \begin{bmatrix} \underline{\Phi}_k^{(k)} & \underline{\Delta}_k^{(k)} \\ \underline{\Psi}_k^{(k)} & \underline{\Theta}_k^{(k)} \end{bmatrix} \right)'$$

$$= Q'_k.$$

PROOF of Theorem 7

From (5),

$$\begin{bmatrix} \widehat{\Phi}_k^{(k)} \\ \widehat{\Psi}_k^{(k)} \end{bmatrix} = - \begin{bmatrix} \widehat{U}_k & \widehat{V}_k \\ \widehat{V}'_k & \widehat{F}_k \end{bmatrix}^{-1} \begin{bmatrix} \widehat{\Gamma}_k \\ \widehat{\Pi}_{-k} \end{bmatrix}.$$

Thus

$$\begin{bmatrix} \widehat{\Phi}_k^{(k)} \\ \widehat{\Psi}_k^{(k)} \end{bmatrix} - \begin{bmatrix} \Phi_k^{(k)} \\ \Psi_k^{(k)} \end{bmatrix} = - \begin{bmatrix} \widehat{U}_k & \widehat{V}_k \\ \widehat{V}_k' & \widehat{F}_k \end{bmatrix}^{-1} \left\{ \begin{bmatrix} \widehat{\Gamma}_k \\ \widehat{\Pi}_{-k} \end{bmatrix} + \begin{bmatrix} \widehat{U}_k & \widehat{V}_k \\ \widehat{V}_k' & \widehat{F}_k \end{bmatrix} \begin{bmatrix} \Phi_k^{(k)} \\ \Psi_k^{(k)} \end{bmatrix} \right\}.$$

Now the jth component of

$$\widehat{\Gamma}_k + \begin{bmatrix} \widehat{U}_k & \widehat{V}_k \end{bmatrix} \begin{bmatrix} \Phi_k^{(k)} \\ \Psi_k^{(k)} \end{bmatrix}$$

is

$$\begin{aligned} & T^{-1} \sum_{t=1}^{T-j} \mathbf{X}'(t) \mathbf{X}(t-j) + \sum_{i=1}^{k} \phi_i T^{-1} \sum_{t=1}^{T-|i-j|} \mathbf{X}'(t) \mathbf{X}(t+|i-j|) \\ & + \sum_{i=1}^{k} \psi_i T^{-1} \sum_{t=1}^{T-|i-j|} \mathbf{X}'(t) W \mathbf{X}(t+|i-j|) \\ & = T^{-1} \sum_{t=1}^{T} \left\{ \mathbf{X}(t) + \sum_{i=1}^{k} \phi_i X(t-i) + \sum_{i=1}^{k} \psi_i W \mathbf{X}(t-i) \right\}' \mathbf{X}(t-j) + o\left(T^{-1+\delta}\right), \end{aligned}$$

almost surely, for any $\delta > 0$, since the difference involves only a finite number of quadratic terms in the components of $\mathbf{X}(t), \mathbf{X}(t-1), \ldots$. Thus the jth component is

$$T^{-1} \sum_{t=1}^{T} \varepsilon(t)' \mathbf{X}(t-j) + o\left(T^{-1+\delta}\right),$$

almost surely as $T \to \infty$. Similarly, the jth component of

$$\widehat{\Pi}_{-k} + \begin{bmatrix} \widehat{V}_{-k} & \widehat{F}_k \end{bmatrix} \begin{bmatrix} \Phi_k^{(k)} \\ \Psi_k^{(k)} \end{bmatrix}$$

is

$$T^{-1} \sum_{t=1}^{T} \varepsilon(t)' W \mathbf{X}(t-j) + o\left(T^{-1+\delta}\right).$$

But both

$$T^{-1} \sum_{t=1}^{T} \varepsilon(t)' \mathbf{X}(t-j)$$

and

$$T^{-1}\sum_{t=1}^{T}\varepsilon\left(t\right)'W\mathbf{X}\left(t-j\right)$$

converge almost surely to 0 by the ergodic theorem. Also, since

$$\begin{bmatrix} \widehat{U}_k & \widehat{V}_k \\ \widehat{V}_k' & \widehat{F}_k \end{bmatrix}$$

converges almost surely to

$$\begin{bmatrix} U_k & V_k \\ V_k' & F_k \end{bmatrix},$$

again by the ergodic theorem, the first part of the theorem follows. Now, letting a and b be arbitrary N-dimensional vectors, we have from the above,

$$\begin{bmatrix} a' & b' \end{bmatrix}\left\{\begin{bmatrix} \widehat{\Gamma}_k \\ \widehat{\Pi}_{-k} \end{bmatrix}+\begin{bmatrix} \widehat{U}_k & \widehat{V}_k \\ \widehat{V}_k' & \widehat{F}_k \end{bmatrix}\begin{bmatrix} \Phi_k^{(k)} \\ \Psi_k^{(k)} \end{bmatrix}\right\}$$
$$=T^{-1}\sum_{t=1}^{T}\varepsilon\left(t\right)'\sum_{j=1}^{k}\left(a_jI_N+b_jW\right)\mathbf{X}\left(t-j\right)+o\left(T^{-1+\delta}\right).$$

But, letting

$$\xi\left(t\right)=\varepsilon\left(t\right)'\sum_{j=1}^{k}\left(a_jI_N+b_jW\right)\mathbf{X}\left(t-j\right),$$

we have

$$E\left\{\xi\left(t\right)\middle|\mathcal{F}_{t-1}\right\}=0$$
$$E\left\{\xi^2\left(t\right)\middle|\mathcal{F}_{t-1}\right\}=\sigma^2\left\{\sum_{i=1}^{k}\mathbf{X}'\left(t-i\right)\left(a_iI_N+b_iW'\right)\sum_{j=1}^{k}\left(a_jI_N+b_jW\right)\mathbf{X}\left(t-j\right)\right\}$$

and

$$E\left\{\xi^2(t)\right\} = \sigma^2 E\left\{\sum_{i=1}^{k} \mathbf{X}'(t-i)\left(a_i I_N + b_i W'\right)\sum_{j=1}^{k}\left(a_j I_N + b_j W\right)\mathbf{X}(t-j)\right\}$$
$$= \sigma^2 \sum_{i=1}^{k}\sum_{j=1}^{k}\left(a_i a_j \gamma_{i-j} + a_i b_j \pi_{i-j} + b_i a_j \pi_{j-i} + b_i b_j \lambda_{i-j}\right)$$
$$= \sigma^2 \begin{bmatrix} a' & b' \end{bmatrix} \begin{bmatrix} U_k & V_k \\ V_k' & F_k \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}.$$

Thus, by Billingsley's martingale central limit theorem (Billingsley (1961)), it follows that

$$\sqrt{T}\left[\ \left\{\widehat{\Phi}_k^{(k)} - \Phi_k^{(k)}\right\}' \quad \left\{\widehat{\Psi}_k^{(k)} - \Psi_k^{(k)}\right\}'\ \right]'$$

is asymptotically normal with mean zero and covariance matrix

$$\sigma^2 \begin{bmatrix} U_k & V_k \\ V_k' & F_k \end{bmatrix}^{-1} \begin{bmatrix} U_k & V_k \\ V_k' & F_k \end{bmatrix} \begin{bmatrix} U_k & V_k \\ V_k' & F_k \end{bmatrix}^{-1}$$
$$= \sigma^2 \begin{bmatrix} U_k & V_k \\ V_k' & F_k \end{bmatrix}^{-1}.$$

PROOF of Theorem 8 The derivations of conditions for weak and strong consistency of the minimiser of $\phi_f(k)$ each depend on the asymptotic properties of

$$\phi_f(k+1) - \phi_f(k) = NT \log \frac{\widehat{\sigma}_{k+1}^2}{\widehat{\sigma}_k^2} + 2f(T).$$

Now,

$$\frac{\widehat{\sigma}_{k+1}^2}{\widehat{\sigma}_k^2} = 1 + \frac{\widehat{\sigma}_{k+1}^2 - \widehat{\sigma}_k^2}{\widehat{\sigma}_k^2}$$
$$= 1 + \frac{\widehat{s}_{k+1} - \widehat{s}_k}{\widehat{s}_k},$$

where $\widehat{s}_k$ denotes the $(1,1)$ element of $\widehat{S}_k$. From (12),

$$
\begin{aligned}
S_{k+1} - S_k &= -S_k \begin{bmatrix} \underline{\phi}_{k+1}^{(k+1)} & \underline{\delta}_{k+1}^{(k+1)} \\ \underline{\psi}_{k+1}^{(k+1)} & \underline{\theta}_{k+1}^{(k+1)} \end{bmatrix} \begin{bmatrix} \phi_{k+1}^{(k+1)} & \delta_{k+1}^{(k+1)} \\ \psi_{k+1}^{(k+1)} & \theta_{k+1}^{(k+1)} \end{bmatrix} \\
&= Q_k \begin{bmatrix} \phi_{k+1}^{(k+1)} & \delta_{k+1}^{(k+1)} \\ \psi_{k+1}^{(k+1)} & \theta_{k+1}^{(k+1)} \end{bmatrix} \\
&= -Q_k T_k^{-1} P_k \\
&= -P_k' T_k^{-1} P_k,
\end{aligned}
$$

by Lemma 6. Thus,

$$
\begin{aligned}
\widehat{s}_{k+1} - \widehat{s}_k &= -\widehat{P}_{k1}' \widehat{T}_k^{-1} \widehat{P}_{k1} \\
&= -\begin{bmatrix} \widehat{\phi}_{k+1}^{(k+1)} & \widehat{\psi}_{k+1}^{(k+1)} \end{bmatrix} \widehat{T}_k \begin{bmatrix} \widehat{\phi}_{k+1}^{(k+1)} \\ \widehat{\psi}_{k+1}^{(k+1)} \end{bmatrix}
\end{aligned}
$$

It follows from Theorem 7 that $\widehat{T}_k$ converges almost surely to T_k, which is positive definite, and that $\widehat{\phi}_{k+1}^{(k+1)}$ and $\widehat{\psi}_{k+1}^{(k+1)}$ converge almost surely to $\phi_{k+1}^{(k+1)}$ and $\psi_{k+1}^{(k+1)}$, which are both zero when $k \geq k_0$, but not when $k = k_0 - 1$. Thus, when $k \geq k_0$, we have

$$
\widehat{s}_{k+1} - \widehat{s}_k = -T^{-1} \sigma^2 z_k' z_k \left\{1 + o(1)\right\},
$$

almost surely as $T \to \infty$, where z_k is defined in Lemma 10 and

$$
\begin{aligned}
\phi_f(k+1) - \phi_f(k) &= NT \log\left(1 + \frac{\widehat{s}_{k+1} - \widehat{s}_k}{\widehat{s}_k}\right) + 2f(T) \\
&= NT \log\left[1 - \frac{T^{-1}\sigma^2 z_k' z_k}{N\sigma^2}\left\{1 + o(1)\right\}\right] + 2f(T) \\
&= -z_k' z_k \left\{1 + o(1)\right\} + 2f(T).
\end{aligned}
$$

Also, when $k < k_0$, we have, almost surely as $T \to \infty$,

$$\widehat{s}_{k+1} - \widehat{s}_k \to \begin{cases} -\begin{bmatrix} \phi_{k+1} & \psi_{k+1} \end{bmatrix} T_k \begin{bmatrix} \phi_{k+1} \\ \psi_{k+1} \end{bmatrix} \leq 0 & ; \quad k < k_0 - 1 \\ -\begin{bmatrix} \phi_{k_0} & \psi_{k_0} \end{bmatrix} T_{k_0-1} \begin{bmatrix} \phi_{k_0} \\ \psi_{k_0} \end{bmatrix} < 0 & ; \quad k = k_0 - 1. \end{cases}$$

It follows that

$$T^{-1} \left\{ \phi_f(k+1) - \phi_f(k) \right\}$$

converges almost surely to

$$N \log \left(1 - s_{k+1}^{-1} \begin{bmatrix} \phi_{k+1} & \psi_{k+1} \end{bmatrix} T_k \begin{bmatrix} \phi_{k+1} \\ \psi_{k+1} \end{bmatrix} \right) \leq 0,$$

if $k < k_0 - 1$, to

$$N \log \left(1 - s_{k_0}^{-1} \begin{bmatrix} \phi_{k_0} & \psi_{k_0} \end{bmatrix} T_{k_0-1} \begin{bmatrix} \phi_{k_0} \\ \psi_{k_0} \end{bmatrix} \right) < 0,$$

if $k = k_0 - 1$, and is asymptotically equivalent to

$$T^{-1} \left\{ -z_k' z_k + 2f(T) \right\} \tag{23}$$

if $k \geq k_0$.

Hence $\phi_f(k)$ is asymptotically non-increasing when $k < k_0 - 1$, and strictly decreasing at $k = k_0 - 1$. Since by Lemma 10 the $z_k' z_k$ are asymptotically independent and χ_2^2 when $k \geq k_0$, $\phi_f(k+1) - \phi_f(k)$ is asymptotically positive (in probability) if $f(T)$ diverges to ∞ with T. Thus, $\widehat{k}$ will be weakly consistent if $f(T)$ diverges to ∞. Of course, consistency is an asymptotic concept, and criteria with larger values of $f(T)$ yield estimators which are equal or smaller. Thus in small samples, we may in fact underestimate the order. It is of interest therefore to choose f so as to fix the (asymptotic) probability of overestimation at some acceptable level. We proceed as in Quinn (1988). Put $f(T) = c$ and note from above that

$$\Pr\left\{\widehat{k} < k_0\right\} \to 0.$$

Then, noting that the $z_k' z_k - 2c$ are asymptotically independent and distributed as $\chi_2^2 - 2c$ for $k = k_0, \ldots, K-1$, we have

$$\lim_{T\to\infty} \Pr\left\{\widehat{k} = k_0 + j\right\}$$
$$= \begin{cases} \Pr\{W_1 < 0, \ldots, W_{K-k_0} < 0\} & ; \quad j = 0 \\ \Pr\{0 < W_j, W_1 < W_j, \ldots, W_{j-1} < W_j, W_{j+1} < W_j, \ldots, W_{K-k_0} < W_j\} & ; \quad j \geq 1, \end{cases}$$

where

$$W_j = \sum_{i=1}^{j} U_i$$

and the U_i are independent and identically distributed with $U_i + 2c$ distributed as χ_2^2. Now

$$\begin{aligned}
&\Pr\{0 < W_j, W_1 < W_j, \ldots, W_{j-1} < W_j, W_{j+1} < W_j, \ldots, W_{K-k_0} < W_j\} \\
&= \Pr\{U_1 + \cdots + U_j > 0, \ldots, U_j > 0, U_{j+1} < 0, \ldots, U_{j+1} + \cdots + U_{K-k_0} < 0\} \\
&= \Pr\{U_1 + \cdots + U_j > 0, \ldots, U_j > 0\} \Pr\{U_{j+1} < 0, \ldots, U_{j+1} + \cdots + U_{K-k_0} < 0\} \\
&= \Pr\{W_1 > 0, \ldots, W_j > 0\} \Pr\{W_1 < 0, \ldots, W_{K-k_0-j} < 0\} \\
&= p_j q_{K-k_0-j},
\end{aligned}$$

where

$$p_j = \Pr\{W_1 > 0, \ldots, W_j > 0\}$$

and

$$q_j = \Pr\{W_1 < 0, \ldots, W_j < 0\}.$$

Note that in the above we have relabelled the U_i in order to obtain expressions in terms of probabilities involving the W_i. Setting $p_0 = q_0 = 1$, and using Sparre-Andersen's theorems (see, for example, Sparre-Andersen(1954), Spitzer(1956) and Feller (1971)), we have

$$\lim_{T\to\infty} \Pr\left\{\widehat{k} = k_0 + j\right\} = p_j q_{K-k_0-j},$$

where the p_i and q_i may be obtained from the generating functions

$$\sum_{i=0}^{\infty} p_i z^i = e^{\sum_{j=1}^{\infty} j^{-1} z^j \Pr\left\{\chi^2_{2j} > 2cj\right\}}$$

and

$$\sum_{i=0}^{\infty} q_i z^i = e^{\sum_{j=1}^{\infty} j^{-1} z^j \Pr\left\{\chi^2_{2j} < 2cj\right\}}.$$

When K is moderately large, we can approximate $\lim_{T\to\infty} \Pr\left\{\widehat{k} = k_0 + j\right\}$ by $p_j q$, where $q = \lim_{k\to\infty} q_k$. Let $\tau_j = q_{j-1} - q_j$. Then

$$\begin{aligned}\sum_{j=1}^{k} \tau_j z^j &= \sum_{j=1}^{k} \left(q_{j-1} - q_j\right) z^j \\ &= 1 + z \sum_{j=0}^{k-1} q_j z^j - \sum_{j=0}^{k} q_j z^j.\end{aligned}$$

Hence

$$\begin{aligned}
1-q &= \lim_{k\to\infty}\left(1-q_k\right) \\
&= \lim_{k\to\infty}\lim_{z\to1^-}\sum_{j=1}^{k}\left(q_{j-1}-q_j\right)z^j \\
&= \lim_{z\to1^-}\lim_{k\to\infty}\sum_{j=1}^{k}\left(q_{j-1}-q_j\right)z^j \\
&= \lim_{z\to1^-}\sum_{j=1}^{\infty}\tau_j z^j \\
&= 1+\lim_{z\to1^-}\left(z\sum_{j=0}^{\infty}q_j z^j-\sum_{j=0}^{\infty}q_j z^j\right) \\
&= 1-\lim_{z\to1^-}\left\{\left(1-z\right)e^{\sum_{j=1}^{\infty}j^{-1}z^j\Pr\left\{\chi^2_{2j}<2cj\right\}}\right\} \\
&= 1-\lim_{z\to1^-}e^{\log(1-z)+\sum_{j=1}^{\infty}j^{-1}z^j\Pr\left\{\chi^2_{2j}<2cj\right\}} \\
&= 1-\lim_{z\to1^-}e^{\sum_{j=1}^{\infty}j^{-1}z^j\left[\Pr\left\{\chi^2_{2j}<2cj\right\}-1\right]} \\
&= 1-e^{\sum_{j=1}^{\infty}-j^{-1}\Pr\left\{\chi^2_{2j}>2cj\right\}}.
\end{aligned}$$

Thus

$$q=e^{-\sum_{j=1}^{\infty}j^{-1}\Pr\left\{\chi^2_{2j}>2cj\right\}}.$$

Note that when $c>1$, $q>0$, since by the central limit theorem as $j\to\infty$,

$$\begin{aligned}
\Pr\left\{\chi^2_{2j}>2cj\right\} &= \Pr\left\{\frac{\chi^2_{2j}-2j}{\sqrt{4j}}>\left(c-1\right)\sqrt{j}\right\} \\
&\sim \Pr\left\{Z>\left(c-1\right)\sqrt{j}\right\} \\
&\sim \frac{1}{\sqrt{2\pi j}}e^{-\frac{1}{2}(c-1)j}.
\end{aligned}$$

Finally, we establish the conditions for strong consistency. From (23), the minimiser of ϕ_f depends on the behaviour of the terms

$$-z_k' z_k+2f\left(T\right),$$

$k = k_0, \ldots, K-1$. From Lemma 10, we have for $k = k_0, \ldots, K-1$,

$$\limsup_T \frac{z_k' z_k}{2 \log \log T} = 1.$$

If

$$\liminf_T \frac{f(T)}{\log \log T} > 1,$$

then

$$\liminf_T \frac{-z_k' z_k + 2f(T)}{2 \log \log T} > 0,$$

and consequently $\widehat{k} \to k_0$, almost surely. If, on the other hand,

$$\liminf_T \frac{f(T)}{\log \log T} < 1,$$

then we have

$$\liminf_T \frac{-z_k' z_k + 2f(T)}{2 \log \log T} < 0,$$

and so the event

$$\phi_f(k+1) - \phi_f(k) < 0$$

occurs infinitely often, for each $k \geq k_0$. Hence $\widehat{k}$ will not converge almost surely to k_0. The condition

$$\liminf_T \frac{f(T)}{\log \log T} \geq 1$$

is thus necessary for strong consistency.

6 Empirical analysis of the properties of $\hat{k}$

Let $\phi_f(k) = NT \log \widehat{\sigma}_k^2 + 2kf(T)$ be the criterion for selecting the order of the space-time AR process. AIC, BIC and the Hannan & Quinn Criterion (HQIC) are obtained by putting $f(T) = 2, \log T$ and $2 \log(\log T)$ (Quinn (1980)), respectively. In order to compare the performance of the three criteria, the procedures developed in the previous sections were implemented in MATLAB and tested on simulated data. The performance is compared only for relatively small values of T. All the simulations reported were performed using pseudo-normal random numbers.

Data was simulated from a system of nine sites with weighting matrix:

$$W = \begin{bmatrix} 0 & .2192 & .1109 & .1173 & .1485 & .1101 & .0928 & .1197 & .0816 \\ .1674 & 0 & .1661 & .0958 & .1352 & .1016 & .0852 & .1614 & .0873 \\ .0759 & .1488 & 0 & .0815 & .1054 & .0981 & .0890 & .2688 & .1325 \\ .0601 & .0642 & .0610 & 0 & .2201 & .2701 & .1665 & .0869 & .0710 \\ .0753 & .0898 & .0782 & .2180 & 0 & .2156 & .1252 & .1219 & .0759 \\ .0470 & .0568 & .0612 & .2252 & .1816 & 0 & .2514 & .0964 & .0804 \\ .0489 & .0588 & .0686 & .1714 & .1302 & .3103 & 0 & .1006 & .1113 \\ .0666 & .1175 & .2185 & .0944 & .1336 & .1256 & .1061 & 0 & .1377 \\ .0616 & .0863 & .1462 & .1047 & .1130 & .1421 & .1593 & .1869 & 0 \end{bmatrix}$$

and the errors were pseudo-$N(0,1)$. Each table below shows the results of 100 replications of each model for varying sample sizes T and different parameters. Tables 1 and 2 show the frequencies of the orders estimated by each criterion for space-time AR processes of order 1. Results for Space-Time AR processes of order 2 are shown in Tables 3 and 4.

From the following tables it can be seen that, for small Φ and Ψ and small T, AIC overestimates the order of the process more often than the other two criteria and BIC underestimates the order more often. HQIC lies between these two criteria by underestimating to a lesser degree than BIC and overestimating to a lesser degree than AIC. On the basis of this we prefer. Negative values of the parameters do not seem to give different results than positive values. When considering different values of the parameters, the correct order seems to be found more often when the value of Φ is larger than when the value of Ψ is larger. This can be observed in sub tables 3, 4 and 5 of Table 2. This observation seems to indicate a stronger weight of Φ in the model and reflects the fact that the influence of Ψ is somewhat weakened by the existence of the weighting matrix W. Although the performance of AIC improves, as T increases, for small Φ and Ψ, this improvement stops after a certain value of T. Also, for larger values of the coefficients, the performance does not appear to improve with T. This is easily explained by the inconsistency of AIC, shown in Theorem 8. Particularly bad performance from all criteria is observed in Table 3 for $\Phi = \Psi = [0.5\ 0.5]'$. The

$\Phi = \Psi = [0.1]$

	T=50				T=100				T=150				T=500			
$\widehat{k}$	0	1	2	>2	0	1	2	>2	0	1	2	>2	0	1	2	>2
AIC	31	52	10	7	4	75	3	3	1	82	9	8	0	81	12	7
HQIC	48	48	1	3	18	79	3	0	2	94	3	1	0	95	4	1
BIC	74	25	1	0	42	57	1	0	13	87	0	0	0	100	0	0

$\Phi = \Psi = [0.2]$

	T=50				T=100				T=150				T=500			
$\widehat{k}$	0	1	2	>2	0	1	2	>2	0	1	2	>2	0	1	2	>2
AIC	0	81	10	9	0	83	8	9	0	82	13	5	0	80	12	8
HQIC	1	92	5	2	0	97	3	0	0	95	4	1	0	94	4	2
BIC	3	96	1	0	0	100	0	0	0	100	0	0	0	100	0	0

$\Phi = \Psi = [0.3]$

	T=50				T=100				T=150				T=500			
$\widehat{k}$	0	1	2	>2	0	1	2	>2	0	1	2	>2	0	1	2	>2
AIC	0	85	10	5	0	87	9	4	0	83	11	6	0	81	11	8
HQIC	0	95	4	1	0	99	1	0	0	97	3	0	0	99	1	0
BIC	0	97	3	0	0	100	0	0	0	99	1	0	0	100	0	0

$\Phi = \Psi = [-0.1]$

	T=50				T=100				T=150				T=500			
$\widehat{k}$	0	1	2	>2	0	1	2	>2	0	1	2	>2	0	1	2	>2
AIC	38	49	8	5	11	68	9	12	0	83	7	10	0	82	10	8
HQIC	53	43	4	0	17	79	2	2	2	95	2	1	0	98	1	1
BIC	64	36	0	0	36	63	1	0	13	86	1	0	0	99	1	0

$\Phi = \Psi = [-0.2]$

	T=50				T=100				T=150				T=500			
$\widehat{k}$	0	1	2	>2	0	1	2	>2	0	1	2	>2	0	1	2	>2
AIC	0	88	10	2	0	82	11	7	0	80	11	9	0	83	8	9
HQIC	2	94	4	0	0	98	1	1	0	96	4	0	0	100	0	0
BIC	6	93	1	0	0	100	0	0	0	100	0	0	0	100	0	0

Table 1: Frequency of correct order selection, for simulated data from $Space - Time\ AR$ models of order 1.

$\Phi = [0.2]\ \Psi = [-0.2]$

	T=50				T=100				T=150				T=500			
$\widehat{k}$	0	1	2	>2	0	1	2	>2	0	1	2	>2	0	1	2	>2
AIC	1	87	5	7	0	83	8	9	0	84	11	5	0	75	16	9
HQIC	3	94	2	1	0	97	2	1	0	96	4	0	0	95	5	0
BIC	7	92	1	0	0	99	1	0	0	100	0	0	0	100	0	0

$\Phi = [-0.2]\ \Psi = [0.2]$

	T=50				T=100				T=150				T=500			
$\widehat{k}$	0	1	2	>2	0	1	2	>2	0	1	2	>2	0	1	2	>2
AIC	0	85	12	3	0	78	11	11	0	79	11	10	0	86	3	11
HQIC	0	93	7	0	0	94	6	0	0	97	2	1	0	97	3	0
BIC	3	95	2	0	0	97	3	0	0	100	0	0	0	100	0	0

$\Phi = [0.3]\ \Psi = [-0.1]$

	T=50				T=100				T=150				T=500			
$\widehat{k}$	0	1	2	>2	0	1	2	>2	0	1	2	>2	0	1	2	>2
AIC	0	85	9	6	0	82	13	5	0	82	12	6	0	79	11	10
HQIC	0	95	3	2	0	95	5	0	0	97	3	0	0	96	2	2
BIC	0	99	1	0	0	98	2	0	0	98	2	0	0	100	0	0

$\Phi = [0.1]\ \Psi = [-0.3]$

	T=50				T=100				T=150				T=500			
$\widehat{k}$	0	1	2	>2	0	1	2	>2	0	1	2	>2	0	1	2	>2
AIC	4	79	9	8	0	80	15	5	0	80	11	9	0	81	12	7
HQIC	10	85	3	2	1	93	6	0	1	94	5	0	0	97	2	1
BIC	24	76	0	0	6	94	0	0	1	99	0	0	0	100	0	0

$\Phi = [0.1]\ \Psi = [0.3]$

	T=50				T=100				T=150				T=500			
$\widehat{k}$	0	1	2	>2	0	1	2	>2	0	1	2	>2	0	1	2	>2
AIC	7	79	8	6	1	84	9	6	0	87	7	6	0	82	12	6
HQIC	18	76	6	0	1	91	5	2	0	96	4	0	0	98	2	0
BIC	29	68	3	0	4	95	1	0	0	100	0	0	0	100	0	0

Table 2: Frequency of correct order selection, for simulated data from $Space - Time\ AR$ models of order 1.

$\Phi = \Psi = [0.1\ 0.1]$

	T=50				T=100				T=150				T=500			
$\widehat{k}$	0	1	2	>2	0	1	2	>2	0	1	2	>2	0	1	2	>2
AIC	25	19	40	16	2	3	80	15	0	3	81	16	0	0	86	14
HQIC	43	15	31	11	9	9	81	1	1	9	86	4	0	0	99	1
BIC	62	13	24	1	36	15	49	0	5	21	74	0	0	0	100	0

$\Phi = \Psi = [0.2\ 0.2]$

	T=50				T=100				T=150				T=500			
$\widehat{k}$	0	1	2	>2	0	1	2	>2	0	1	2	>2	0	1	2	>2
AIC	0	1	84	15	0	0	83	17	0	0	81	19	0	0	90	10
HQIC	0	1	96	3	0	0	97	3	0	0	95	5	0	0	100	0
BIC	0	3	96	1	0	0	99	1	0	0	100	0	0	0	100	0

$\Phi = \Psi = [0.3\ 0.3]$

	T=50				T=100				150				T=500			
$\widehat{k}$	0	1	2	>2	0	1	2	>2	0	1	2	>2	0	1	2	>2
AIC	0	0	83	17	0	0	79	21	0	0	89	11	0	0	79	21
HQIC	0	0	96	4	0	0	97	3	0	0	96	4	0	0	96	4
BIC	0	0	98	2	0	0	100	0	0	0	100	0	0	0	100	0

$\Phi = \Psi = [0.5\ 0.5]$

	T=50				T=100				T=150				T=500			
$\widehat{k}$	0	1	2	>2	0	1	2	>2	0	1	2	>2	0	1	2	>2
AIC	0	0	36	57	0	0	26	74	0	0	19	81	0	0	20	80
HQIC	0	0	45	55	0	0	32	68	0	0	34	66	0	0	31	69
BIC	0	0	54	46	0	0	40	60	0	0	42	58	0	0	37	63

Table 3: Frequency of correct order selection, for simulated data from $Space - Time\ AR$ models of order 2.

$\Phi = [0.5\ 0.1]$ $\Psi = [0.5\ 0.1]$

	T=50				T=100				T=150				T=500			
$\widehat{k}$	0	1	2	>2	0	1	2	>2	0	1	2	>2	0	1	2	>2
AIC	0	33	54	13	0	5	77	18	0	3	85	12	0	0	83	17
HQIC	0	51	45	4	0	20	76	4	0	6	91	3	0	0	97	3
BIC	0	64	35	1	0	39	59	2	0	14	86	0	0	0	100	0

$\Phi = [0.5\ 0.1]$ $\Psi = [0.1\ 0.5]$

	T=50				T=100				T=150				T=500			
$\widehat{k}$	0	1	2	>2	0	1	2	>2	0	1	2	>2	0	1	2	>2
AIC	0	0	88	12	0	0	85	16	0	0	81	19	0	0	79	21
HQIC	0	0	97	3	0	0	93	7	0	0	92	8	0	0	97	3
BIC	0	4	96	0	0	0	98	2	0	0	99	1	0	0	100	0

$\Phi = [0.1\ 0.3]$ $\Psi = [-0.2\ -0.2]$

	T=50				T=100				T=150				T=500			
$\widehat{k}$	0	1	2	>2	0	1	2	>2	0	1	2	>2	0	1	2	>2
AIC	0	0	80	20	0	0	87	13	0	0	86	14	0	0	82	18
HQIC	0	0	90	10	0	0	99	1	0	0	94	6	0	0	98	2
BIC	0	0	95	5	0	0	100	0	0	0	99	1	0	0	100	0

Table 4: Frequency of correct order selection, for simulated data from $Space - Time\ AR$ models of order 2.

reason for this is that for these values, 2 of the zeros in (4) are 0. (*Something wrong here. Two zeros appear to be outside the unit circle. Could you do the simulations again with 0.45 instead of 0.5?)*

7 Data Analysis: Space-time AR models for Carbon monoxide concentrations in Venice

The atmospheric concentration of carbon monoxide (CO) is a central issue in environmental monitoring, because of the direct relationship between excess CO and global warming (greenhouse effect). The problem is spatio-temporal in nature and given time series from different monitoring stations, we expect better insights into the dynamics of the process by fitting a spatio-temporal model to the data rather than analysing the spatial and temporal features separately.

The data consist of 300 hourly carbon monoxide (CO) concentrations (in micrograms per cubic meter) recorded in September 1995 at four different locations in Venice (see Figure 1). The data were considered in Tonellato(2001) and are available at http://www.blackwellpublishers.co.uk/rss/datasets/tonellato.ZIP. In Tonellato(2001) a multivariate time series model is fitted to the data using Bayesian methods. Tonellato assumes that the process is isotropic, which may be difficult to justify. With data available at only four locations it is also difficult to model the spatial covariance structure of the system. Hence, we believe that the space-time autoregressive model is suitable for modelling this data.

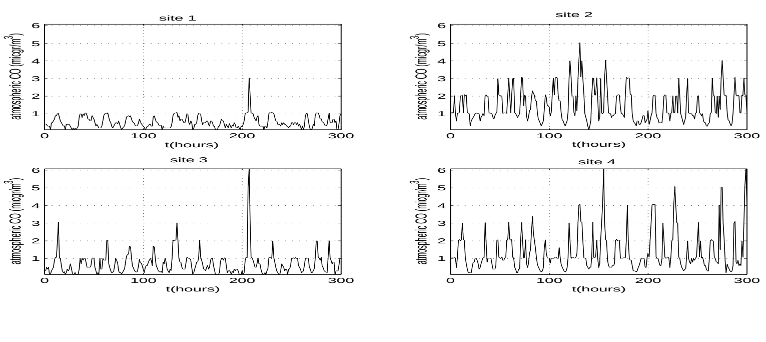

Figure 1: Original series of hourly carbon monoxide in 4 stations in Venice

Site	2	3	4
1	1.680	1.420	1.356
2		0.624	1.176
3			0.672

Table 5: Distances between sites (in km)

7·1 *Analysis*

We follow Tonellato(2001) by replacing the missing data with the means of the values at the same site and same hour over the whole sample period. We analyse the logarithms of the data, depicted in Figure 2. *(This figure seems wrong. Compare with the one I've attached. In particular, compare the first ten values in the first picture. I'm worried that the analysis might not be correct.)* There is evidence that this transformation stabilises the variance but we believe the data is still non-gaussian.

The distances between the sites (Table 5) are used to construct the weighting matrix. The weights are taken as the inverse of the corresponding distances between the sites. The weighting matrix is standardised such that each row sum is 1:

$$W = \begin{bmatrix} 0 & 0.2922 & 0.3457 & 0.3621 \\ 0.1946 & 0 & 0.5274 & 0.2780 \\ 0.1856 & 0.4223 & 0 & 0.3921 \\ 0.2398 & 0.2764 & 0.4838 & 0 \end{bmatrix}$$

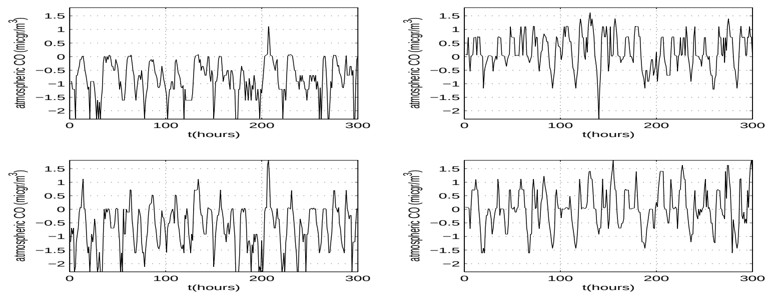

Figure 2: Logarithm of series of hourly carbon monoxide in 4 stations in Venice

Following Tonellato, we subtract the seasonal component present in each time series individually prior to modelling. Assuming that the length of each cycle is 24 hours and denoting by $\mathbf{Y}(t)$ the logarithm of the original observations, we write

$$\mathbf{Y}(t) = S(t)\xi + \mathbf{X}(t) \tag{24}$$

where

$$S(t) = \begin{bmatrix} \cos(\pi t/12) & \cdots & \cos(6\pi t/12) & \sin(\pi t/12) & \cdots & \sin(6\pi t/12) \end{bmatrix} \otimes I_4$$

and

$$\xi = \begin{bmatrix} \xi_{1,1} & \cdots & \xi_{1,6} & \cdots & \xi_{4,1} & \cdots & \xi_{4,6} & \xi^*_{1,1} & \cdots & \xi^*_{1,6} & \cdots & \xi^*_{4,1} & \cdots & \xi^*_{4,6} \end{bmatrix}'$$

and $\mathbf{X}(t)$ is the deseasonalised process in Figure 3, which will be used for fitting a space-time AR model. In practice, $\mathbf{X}(t)$ are the residuals obtained after fitting the model (24) by least squares to the transformed data.

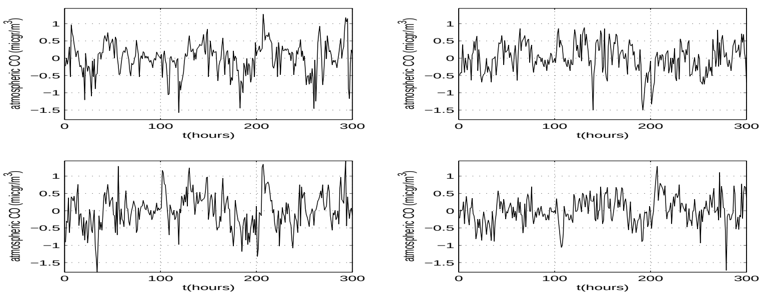

Figure 3: Deseasonalised series(using harmonics) of log-hourly carbon monoxide in 4 stations in Venice.

Using our methodology for order determination and using the recursive estimation techniques, we found that the best order for the space-time AR model is 1 and the corresponding estimated parameters are $\hat{\phi} = -0.5202$, $\hat{\psi} = -0.1173$, $\hat{\sigma}^2 = 0.1341$. The graph of the residuals obtained after fitting the space-time AR model (Figure 4) does not present any specific pattern. The residual autocorrelation functions are not significantly different from zero and the cross correlations are also very close to zero .

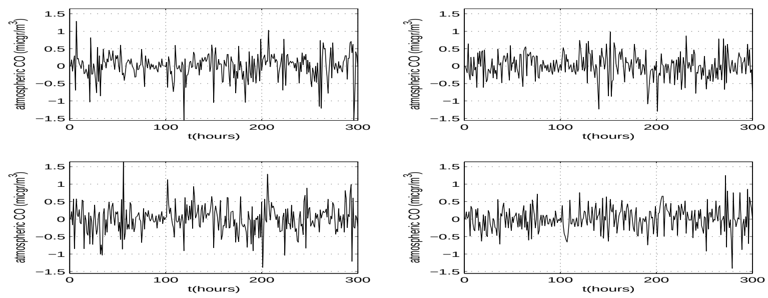

Figure 4: Residual after fitting a space-time AR(1) model to the data in Figure 3.

Acknowledgment

The work of Ana Mónica C.Antunes was supported by a Grant from Fundação para a Ciência e a Tecnologia, Portugal (SFRH/BD/1473/2000) and the author is grateful for the Grant.

Appendix A. Application of the Law of the iterated logarithm

The following lemmas are needed for the proof of Theorem 8.

LEMMA 9 *Let* $\widehat{P}_{k1}$ *denote the first column of* $\widehat{P}_k$*, where* P_k *is defined in Theorem 3. Then, when* $k \geq k_0$*, the true order,*

$$\widehat{P}_{k1} = \begin{bmatrix} T^{-1} \sum_{t=1}^{T} \underline{\varepsilon}_k' \left(t-k-1\right) \varepsilon\left(t\right) \\ T^{-1} \sum_{t=1}^{T} \underline{u}_k' \left(t-k-1\right) \varepsilon\left(t\right) \end{bmatrix} \left\{1 + o\left(1\right)\right\},$$

almost surely as $T \to \infty$.

LEMMA 10 *Let*

$$z_k = T^{1/2} \sigma^{-1} T_k^{-1/2} \widehat{P}_{k1},$$

where $T_k^{-1/2}$ *is any matrix whose square is* T_k^{-1}*. Then*

1. *For* $k = k_0, \ldots, K$*, where* K *is fixed, the* z_k *are asymptotically independent and normally distributed with zero means and identity covariance matrices.*

2. For $k \geq k_0$, if α is any 2-dimensional vector,

$$\limsup_{T} \sqrt{\frac{1}{2\alpha'\alpha \log\log T}} \, |\alpha' z_k| = 1,$$

almost surely.

3. For $k = k_0, \ldots, K$, where K is fixed,

$$\limsup_{T} \frac{z_k' z_k}{2 \log\log T} = 1,$$

almost surely.

PROOF of Lemma 9 From (8) ,

$$\begin{aligned}
\widehat{P}_{k1} &= \begin{bmatrix} \widehat{\gamma}_{k+1} \\ \widehat{\pi}_{-k-1} \end{bmatrix} + \begin{bmatrix} \widetilde{\widehat{\Gamma}}'_k & \widetilde{\widehat{\Pi}}'_k \\ \widetilde{\widehat{\Pi}}'_{-k} & \widetilde{\widehat{\Lambda}}'_k \end{bmatrix} \begin{bmatrix} \widehat{\Phi}_k^{(k)} \\ \widehat{\Psi}_k^{(k)} \end{bmatrix} \\
&= \begin{bmatrix} \widehat{\gamma}_{k+1} \\ \widehat{\pi}_{-k-1} \end{bmatrix} + \begin{bmatrix} \widetilde{\widehat{\Gamma}}'_k & \widetilde{\widehat{\Pi}}'_k \\ \widetilde{\widehat{\Pi}}'_{-k} & \widetilde{\widehat{\Lambda}}'_k \end{bmatrix} \begin{bmatrix} \Phi_k^{(k)} \\ \Psi_k^{(k)} \end{bmatrix} \\
&+ \begin{bmatrix} \widetilde{\widehat{\Gamma}}'_k & \widetilde{\widehat{\Pi}}'_k \\ \widetilde{\widehat{\Pi}}'_{-k} & \widetilde{\widehat{\Lambda}}'_k \end{bmatrix} \left(\begin{bmatrix} \widehat{\Phi}_k^{(k)} \\ \widehat{\Psi}_k^{(k)} \end{bmatrix} - \begin{bmatrix} \Phi_k^{(k)} \\ \Psi_k^{(k)} \end{bmatrix} \right).
\end{aligned}$$

When $k \geq k_0$,

$$\begin{bmatrix} \widehat{\gamma}_{k+1} \\ \widehat{\pi}_{-k-1} \end{bmatrix} + \begin{bmatrix} \widehat{\widetilde{\Gamma}}'_k & \widehat{\widetilde{\Pi}}'_k \\ \widehat{\widetilde{\Pi}}'_{-k} & \widehat{\widetilde{\Lambda}}'_k \end{bmatrix} \begin{bmatrix} \Phi_k^{(k)} \\ \Psi_k^{(k)} \end{bmatrix}$$
$$= \begin{bmatrix} T^{-1}\sum_{t=1}^{T} \mathbf{X}'(t)\,\mathbf{X}(t+k+1) \\ T^{-1}\sum_{t=1}^{T} \mathbf{X}'(t)\,W'\mathbf{X}(t+k+1) \end{bmatrix}$$
$$+ \sum_{j=1}^{k_0} \phi_j \begin{bmatrix} T^{-1}\sum_{t=1}^{T} \mathbf{X}'(t)\,\mathbf{X}(t+k+1-j) \\ T^{-1}\sum_{t=1}^{T} \mathbf{X}'(t)\,W'\mathbf{X}(t+k+1-j) \end{bmatrix}$$
$$+ \sum_{j=1}^{k_0} \psi_j \begin{bmatrix} T^{-1}\sum_{t=1}^{T} \mathbf{X}'(t)\,W'\mathbf{X}(t+k+1-j) \\ T^{-1}\sum_{t=1}^{T} \mathbf{X}'(t)\,W'W\mathbf{X}(t+k+1-j) \end{bmatrix}$$
$$= T^{-1} \begin{bmatrix} \sum_{t=1}^{T} \mathbf{X}'(t)\,\varepsilon(t+k+1) \\ \sum_{t=1}^{T} \mathbf{X}'(t)\,W'\varepsilon(t+k+1) \end{bmatrix}.$$

Now, from the proof of Theorem 7,

$$\begin{bmatrix} \widehat{\widetilde{\Gamma}}'_k & \widehat{\widetilde{\Pi}}'_k \\ \widehat{\widetilde{\Pi}}'_{-k} & \widehat{\widetilde{\Lambda}}'_k \end{bmatrix} \left(\begin{bmatrix} \widehat{\Phi}_k^{(k)} \\ \widehat{\Psi}_k^{(k)} \end{bmatrix} - \begin{bmatrix} \Phi_k^{(k)} \\ \Psi_k^{(k)} \end{bmatrix} \right)$$
$$= - \begin{bmatrix} \widetilde{\Gamma}'_k & \widetilde{\Pi}'_k \\ \widetilde{\Pi}'_{-k} & \widetilde{\Lambda}'_k \end{bmatrix} \begin{bmatrix} U_k & V_k \\ V'_k & F_k \end{bmatrix}^{-1} \left(\begin{bmatrix} \widehat{\Gamma}_k \\ \widehat{\Pi}_{-k} \end{bmatrix} + \begin{bmatrix} \widehat{U}_k & \widehat{V}_k \\ \widehat{V}'_k & \widehat{F}_k \end{bmatrix} \begin{bmatrix} \Phi_k^{(k)} \\ \Psi_k^{(k)} \end{bmatrix} \right) \{1 + o(1)\}$$
$$= - \left(\begin{bmatrix} U_k & V_k \\ V'_k & F_k \end{bmatrix}^{-1} \begin{bmatrix} \widetilde{\Gamma}_k & \widetilde{\Pi}_{-k} \\ \widetilde{\Pi}_k & \widetilde{\Lambda}_k \end{bmatrix} \right)' \left(\begin{bmatrix} \widehat{\Gamma}_k \\ \widehat{\Pi}_{-k} \end{bmatrix} + \begin{bmatrix} \widehat{U}_k & \widehat{V}_k \\ \widehat{V}'_k & \widehat{F}_k \end{bmatrix} \begin{bmatrix} \Phi_k^{(k)} \\ \Psi_k^{(k)} \end{bmatrix} \right) \{1 + o(1)\}$$
$$= - \left(\begin{bmatrix} U_k & V'_k \\ V_k & F_k \end{bmatrix}^{-1} \begin{bmatrix} \Gamma_k & \Pi_{-k} \\ \Pi_k & \Lambda_k \end{bmatrix} \right)' \left(\begin{bmatrix} \widehat{\Gamma}_k \\ \widehat{\Pi}_{-k} \end{bmatrix} + \begin{bmatrix} \widehat{U}_k & \widehat{V}_k \\ \widehat{V}'_k & \widehat{F}_k \end{bmatrix} \begin{bmatrix} \Phi_k^{(k)} \\ \Psi_k^{(k)} \end{bmatrix} \right) \{1 + o(1)\}$$
$$= \begin{bmatrix} \underline{\widetilde{\Phi}}_k^{(k)} & \underline{\widetilde{\Delta}}_k^{(k)} \\ \underline{\widetilde{\Psi}}_k^{(k)} & \underline{\widetilde{\Theta}}_k^{(k)} \end{bmatrix}' \left(\begin{bmatrix} \widehat{\Gamma}_k \\ \widehat{\Pi}_{-k} \end{bmatrix} + \begin{bmatrix} \widehat{U}_k & \widehat{V}_k \\ \widehat{V}'_k & \widehat{F}_k \end{bmatrix} \begin{bmatrix} \Phi_k^{(k)} \\ \Psi_k^{(k)} \end{bmatrix} \right) \{1 + o(1)\}.$$

Thus

$$\widehat{P}_{k1} = \begin{bmatrix} T^{-1}\sum_{t=1}^{T}\mathbf{X}'(t-k-1)\,\varepsilon(t) \\ T^{-1}\sum_{t=1}^{T}\mathbf{X}'(t-k-1)\,W'\varepsilon(t) \end{bmatrix}$$
$$+ \begin{bmatrix} \underline{\widetilde{\Phi}}_k^{(k)} & \underline{\widetilde{\Delta}}_k^{(k)} \\ \underline{\widetilde{\Psi}}_k^{(k)} & \underline{\widetilde{\Theta}}_k^{(k)} \end{bmatrix}' \begin{bmatrix} \left\{T^{-1}\sum_{t=1}^{T}\mathbf{X}'(t-i)\,\varepsilon(t)\right\}_{i=1,\dots,k} \\ \left\{T^{-1}\sum_{t=1}^{T}\mathbf{X}'(t-i)\,W'\varepsilon(t)\right\}_{i=1,\dots,k} \end{bmatrix} \{1+o(1)\}$$
$$= \begin{bmatrix} T^{-1}\sum_{t=1}^{T}\underline{\varepsilon}_k'(t-k-1)\,\varepsilon(t) \\ T^{-1}\sum_{t=1}^{T}\underline{u}_k'(t-k-1)\,\varepsilon(t) \end{bmatrix}\{1+o(1)\}.$$

PROOF of Lemma 10 The asymptotic behaviour of $\widehat{P}_{k1}$ is, from the above, the same as that of

$$\begin{bmatrix} T^{-1}\sum_{t=1}^{T}\underline{\varepsilon}_k'(t-k-1)\,\varepsilon(t) \\ T^{-1}\sum_{t=1}^{T}\underline{u}_k'(t-k-1)\,\varepsilon(t) \end{bmatrix}.$$

Now

$$E\left\{\begin{bmatrix} \underline{\varepsilon}_k'(t-k-1) \\ \underline{u}_k'(t-k-1) \end{bmatrix}\varepsilon(t)\,\varepsilon'(t)\begin{bmatrix} \underline{\varepsilon}_k(t-k-1) & \underline{u}_k(t-k-1) \end{bmatrix}\right\}$$
$$= E\left[E\left\{\begin{bmatrix} \underline{\varepsilon}_k'(t-k-1) \\ \underline{u}_k'(t-k-1) \end{bmatrix}\varepsilon(t)\,\varepsilon'(t)\begin{bmatrix} \underline{\varepsilon}_k(t-k-1) & \underline{u}_k(t-k-1) \end{bmatrix}\middle|\mathcal{F}_{t-1}\right\}\right]$$
$$= \sigma^2 E\left\{\begin{bmatrix} \underline{\varepsilon}_k'(t-k-1) \\ \underline{u}_k'(t-k-1) \end{bmatrix}\begin{bmatrix} \underline{\varepsilon}_k(t-k-1) & \underline{u}_k(t-k-1) \end{bmatrix}\right\}$$
$$= \sigma^2 T_k,$$

by Theorem 4, since both $\underline{\varepsilon}_k'(t-k-1)$ and $\underline{u}_k'(t-k-1)$ are constructed from $\mathbf{X}(t-k-1),\dots,$ $\mathbf{X}(t-1)$ and are thus fixed given $\mathcal{F}_{t-1}$. Again using Billingsley's martingale central limit theorem (Billingsley (1961)), $\alpha' z_k$ is asymptotically normal with zero mean and variance $\alpha'\alpha$. In fact, if α_k, $k = k_0,\dots,K$ are arbitrary 2-dimensional vectors, then $\alpha_k' z_k$, $k = k_0,\dots,K$ are asymptotically independent. To see this, without loss of generality choose k and l with $k > l \geq k_0$. Then

$$E\left\{\left[\begin{array}{c}\sum_{t=1}^{T}\underline{\varepsilon}_k'\left(t-k-1\right)\varepsilon\left(t\right)\\ \sum_{t=1}^{T}\underline{u}_k'\left(t-k-1\right)\varepsilon\left(t\right)\end{array}\right]\left[\begin{array}{cc}\sum_{t=1}^{T}\varepsilon'\left(t\right)\underline{\varepsilon}_l\left(t-l-1\right) & \sum_{t=1}^{T}\varepsilon'\left(t\right)\underline{u}_l\left(t-l-1\right)\end{array}\right]\right\}$$
$$=\sigma^2 E\left[\begin{array}{cc}\sum_{t=1}^{T}\underline{\varepsilon}_k'\left(t-k-1\right)\underline{\varepsilon}_l\left(t-l-1\right) & \sum_{t=1}^{T}\underline{\varepsilon}_k'\left(t-k-1\right)\underline{u}_l\left(t-l-1\right)\\ \sum_{t=1}^{T}\underline{u}_k'\left(t-k-1\right)\underline{\varepsilon}_l\left(t-l-1\right) & \sum_{t=1}^{T}\underline{u}_k'\left(t-k-1\right)\underline{u}_l\left(t-l-1\right)\end{array}\right].$$

Now, using (7) and since, for $i = 0, 1, \ldots, l$

$$1 \leq k - l + i \leq k,$$

we obtain

$$\begin{aligned}
&E\left\{\underline{\varepsilon}_k'\left(t-k-1\right)\underline{\varepsilon}_l\left(t-l-1\right)\right\}\\
&=E\left[\left\{\mathbf{X}\left(t-k-1\right)+\sum_{j=1}^{k}\left(\underline{\phi}_j^{(k)}I_N+\underline{\psi}_j^{(k)}W\right)\mathbf{X}\left(t-k-1+j\right)\right\}'\right.\\
&\left.\left\{\mathbf{X}\left(t-l-1\right)+\sum_{i=1}^{l}\left(\underline{\phi}_i^{(l)}I_N+\underline{\psi}_i^{(l)}W\right)\mathbf{X}\left(t-l-1+i\right)\right\}\right]\\
&=\left\{\gamma_{k-l}+\sum_{j=1}^{k}\left(\underline{\phi}_j^{(k)}\gamma_{k-l-j}+\underline{\psi}_j^{(k)}\pi_{j+l-k}\right)\right\}\\
&+\sum_{i=1}^{l}\underline{\phi}_i^{(l)}\left\{\gamma_{k-l+i}+\sum_{j=1}^{k}\left(\underline{\phi}_j^{(k)}\gamma_{k-l+i-j}+\underline{\psi}_j^{(k)}\pi_{j-i+l-k}\right)\right\}\\
&+\sum_{i=1}^{l}\underline{\psi}_i^{(l)}\left\{\pi_{k-l+i}+\sum_{j=1}^{k}\left(\underline{\phi}_j^{(k)}\pi_{k-l+i-j}+\underline{\psi}_j^{(k)}\lambda_{k-l+i-j}\right)\right\}\\
&=0.
\end{aligned}$$

Similarly,

$$
\begin{aligned}
&E\left\{\underline{u}_k'(t-k-1)\,\underline{u}_l(t-l-1)\right\}\\
&= E\left[\left\{W\mathbf{X}(t-k-1)+\sum_{j=1}^{k}\left(\underline{\delta}_j^{(k)}I_N+\underline{\theta}_j^{(k)}W\right)\mathbf{X}(t-k-1+j)\right\}'\right.\\
&\left.\left\{W\mathbf{X}(t-l-1)+\sum_{i=1}^{l}\left(\underline{\delta}_i^{(l)}I_N+\underline{\theta}_i^{(l)}W\right)\mathbf{X}(t-l-1+i)\right\}\right]\\
&= \left\{\lambda_{k-l}+\sum_{j=1}^{k}\left(\underline{\delta}_j^{(k)}\pi_{k-l-j}+\underline{\theta}_j^{(k)}\lambda_{k-l-j}\right)\right\}\\
&+\sum_{i=1}^{l}\underline{\delta}_i^{(l)}\left\{\pi_{l-k-i}+\sum_{j=1}^{k}\left(\underline{\delta}_j^{(k)}\gamma_{k-l+i-j}+\underline{\theta}_j^{(k)}\pi_{l-k+j-i}\right)\right\}\\
&+\sum_{i=1}^{l}\underline{\theta}_i^{(l)}\left\{\lambda_{k-l+i}+\sum_{j=1}^{k}\left(\underline{\delta}_j^{(k)}\pi_{k-l+i-j}+\underline{\theta}_j^{(k)}\lambda_{k-l+i-j}\right)\right\}\\
&= 0,
\end{aligned}
$$

$$
\begin{aligned}
&E\left\{\underline{\varepsilon}_k'(t-k-1)\,\underline{u}_l(t-l-1)\right\}\\
&= E\left[\left\{\mathbf{X}(t-k-1)+\sum_{j=1}^{k}\left(\underline{\phi}_j^{(k)}I_N+\underline{\psi}_j^{(k)}W\right)\mathbf{X}(t-k-1+j)\right\}'\right.\\
&\left.\left\{W\mathbf{X}(t-l-1)+\sum_{i=1}^{l}\left(\underline{\delta}_i^{(l)}I_N+\underline{\theta}_i^{(l)}W\right)\mathbf{X}(t-l-1+i)\right\}\right]\\
&= \left\{\pi_{k-l}+\sum_{j=1}^{k}\left(\underline{\phi}_j^{(k)}\pi_{k-l-j}+\underline{\psi}_j^{(k)}\lambda_{k-l-j}\right)\right\}\\
&+\sum_{i=1}^{l}\underline{\delta}_i^{(l)}\left\{\gamma_{k-l+i}+\sum_{j=1}^{k}\left(\underline{\phi}_j^{(k)}\gamma_{k-l+i-j}+\underline{\psi}_j^{(k)}\pi_{l-k+j-i}\right)\right\}\\
&+\sum_{i=1}^{l}\underline{\theta}_i^{(l)}\left\{\pi_{k-l+i}+\sum_{j=1}^{k}\left(\underline{\phi}_j^{(k)}\pi_{k-l+i-j}+\underline{\psi}_j^{(k)}\lambda_{k-l+i-j}\right)\right\}\\
&= 0
\end{aligned}
$$

and

$$
\begin{aligned}
&E\left\{\underline{u}_k'\left(t-k-1\right)\underline{\varepsilon}_l\left(t-l-1\right)\right\} \\
&= E\left[\left\{W\mathbf{X}\left(t-k-1\right)+\sum_{j=1}^{k}\left(\underline{\delta}_j^{(k)}I_N+\underline{\theta}_j^{(k)}W\right)\mathbf{X}\left(t-k-1+j\right)\right\}'\right. \\
&\left.\left\{\mathbf{X}\left(t-l-1\right)+\sum_{i=1}^{l}\left(\underline{\phi}_i^{(l)}I_N+\underline{\psi}_i^{(l)}W\right)\mathbf{X}\left(t-l-1+i\right)\right\}\right] \\
&= \left\{\pi_{l-k}+\sum_{j=1}^{k}\left(\underline{\delta}_j^{(k)}\gamma_{l-k+j}+\underline{\theta}_j^{(k)}\pi_{l-k+j}\right)\right\} \\
&+\sum_{i=1}^{l}\underline{\phi}_i^{(l)}\left\{\pi_{l-k-i}+\sum_{j=1}^{k}\left(\underline{\delta}_j^{(k)}\gamma_{l-k+j-i}+\underline{\theta}_j^{(k)}\pi_{l-k+j-i}\right)\right\} \\
&+\sum_{i=1}^{l}\underline{\psi}_i^{(l)}\left\{\lambda_{k-l+i}+\sum_{j=1}^{k}\left(\underline{\delta}_j^{(k)}\pi_{k-l+i-j}+\underline{\theta}_j^{(k)}\lambda_{k-l+i-j}\right)\right\} \\
&= 0.
\end{aligned}
$$

The z_k are thus jointly asymptotically normal and independent.

Using Stout's law of the iterated logarithm (Stout(1970)), we also have

$$\limsup_T \sqrt{\frac{1}{2\alpha'\alpha\log\log T}}\left|\alpha' z_k\right| = 1$$

almost surely. Using the methods in Hannan (1980), which we do not reproduce here, it is also the case that

$$\limsup_T \frac{z_k' z_k}{2\log\log T} = 1.$$

References

Billingsley, P. (1961). The Lindeberg-lévy theorem for martingales. *Proc. Amer. Math. Soc* **12**, 788–792.

Durbin, J. (1960). The fitting of time series models. *Rev. Inst. Statist.* **28**, 233–244.

Epperson, B. (2000). Spatial and space-time correlations in ecological models. *Ecological Modelling* **132**, 63–76.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications, Vol. 2*. New York: Wiley, 2nd ed.

Garrido, R. (2000). Spatial interaction between the truck flows through the Mexico-Texas border. *Transportation Research, Part A* **34**, 23–33.

Guttorp, P., Meiring, W. & Sampson, P. D. (1994). A space-time analysis of ground level ozone data. *Environmetrics* **5**, 241–254.

Hannan, E.J. (1980). The estimation of the order of an ARMA process. *Ann. Statist.* **8**, 1071–1081.

Hannan, E. & Deistler, M. (1988). *The statistical Theory of linear systems*. New York: John Wiley & Sons, 1st ed.

Haslett, J. & Raftery, A. E. (1989). Space-time modelling with long-memory dependence: assessing Ireland's wind power resourse. *Appl. Statist.* **38**, 1–50.

Mardia, K. V., Goodal, C., Redfern, E. J. & Alonso, F. J. (1998). The kriged Kalman filter. *Test* **7**, 217–285.

Pfeifer, P. & Deutsch, S. (1980). A three stage interactive procedure for space-time modeling. *Technometrics* **22**, 35–47.

Quinn, B. (1980). Order determination for a multivariate autoregression. *J. Roy. Statist. Soc. B* **42**, 182–185.

Quinn, B. (1988). A note on AIC order determination for multivariate autoregressions. *J. Time Ser. Anal.* **9**, 241–244.

Shaddick, G. & Wakefield, J. (2002). Modelling daily multivariate pollutant data at multiple sites. *Appl. Statist.* **51**, 351–372.

Sparre-Andersen, E. (1954). On the fluctuations of sums of random variables ii. *Math. Scand.* **2**, 195–223.

Spitzer, F. (1956). A combinatorial lemma and its applications to probability theory.

Trans. Am. Math. Soc. **82**, 323–339.

STOFFER, D. (1986). Estimation and identification of space-time ARMAX models in the presence of missing data. *J. Amer. Statist. Assoc.* **81**, 762–772.

STOUT, W. (1970). The Hartman-Wintner law of the iterated logarithm for martingales. *Ann. Math. Statist.* **41**, 2158–2160.

SUBBA RAO, T. & ANTUNES, A. (2004). Spatio temporal modelling of temperature time series: A comparative study. In *Time series analysis and applications to geophysical systems*, F. Schoenberg, D.R. Brillinger, E. Robinson, eds, vol. 139 of *IMA Volumes in Mathematics and its Applications*. New York: Springer-Verlag,123–150.

TONELLATO, S. F. (2001). A multivariate time series model for the analysis and prediction of carbon monoxide atmospheric concentrations. *Appl. Statist.* **50**, 187–200.

WHITTLE, P. (1963). On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. *Biometrika* **50**, 129–134.

WICKLE, C. K. & CRESSIE, N. (1999). A dimension reduced approach to space-time Kalman filtering. *Biometrika* **86**, 815–829.