Finding common structure in multiple time series via structured priors for autoregressive processes

Gabriel Huerta*† and Raquel Prado‡§

Abstract

We present structured prior modeling in multiple time series, focusing on latent component structure for a collection of autoregressive processes. Similar to the univariate case, the state-space representation of these vector processes implies that each univariate time series can be decomposed into simple underlying components. Such components may have a common structure across the series that define the vector process. The prior specification proposed here extends the class of prior distributions for univariate autoregressions presented in Huerta and West (1999) to a multivariate context. Additionally, this approach allows the consideration of uncertainty on the number of latent processes across the multiple series and consequently, it handles model order uncertainty in the vector autoregressive framework. Posterior inference and implementation are developed via customized Markov chain Monte Carlo (MCMC) methods. Issues related to inference and exploration of the posterior distribution are discussed. Illustrative data analyses are presented.

Keywords: multiple time series, VAR models, model order uncertainty, structured priors, latent components, MCMC methods.

*Department of Mathematics and Statistics, University of New Mexico, Albuquerque, New Mexico 87131-1141, USA. *E-mail:* ghuerta@stat.unm.edu.

†CIMAT, Guanajuato, México.

‡Department of Applied Mathematics and Statistics. Baskin School of Engineering. University of California, Santa Cruz 1156 High Street, Santa Cruz CA 95064, USA. *E-mail:* raquel@ams.ucsc.edu.

§CESMa, Universidad Simón Bolívar, Venezuela.

1 Introduction

The focus of this paper is on developing Bayesian models for the analysis of multivariate time series. Particularly, we propose a prior specification for vector autoregressive (VAR) processes with coefficient matrices that are diagonal. These models are motivated by data arising in areas such signal processing. The data usually consist of multiple signals recorded simultaneously, where each signal has an underlying structure, possibly but not necessarily quasi-periodic, that can be adequately captured using autoregressive (AR) models. Univariate time series arising in applied fields that involve seismic recordings, environmental time series, biomedical signals and speech signals, to mention just a few examples, have such characteristics and have been successfully analyzed in recent years via autoregressive processes, or sophisticated models that involve autoregressive components (Aguilar *et al.*, 1999; Godsill and Rayner, 1998; West *et al.*, 1999; Krystal *et al.*, 1999; Kitagawa and Gersch, 1996).

One of the main features in the development of structured priors for vector autoregressions is that they allow the modeling of uncertainty in the number and form of the latent processes related to each series. Also, they permit the expression of prior beliefs on characteristic AR roots, which may be unitary or zero, while simultaneously allow the inclusion of common latent components across the series, as well as lag-lead structure. Computational difficulties arise when considering many multiple series with a rich latent common structure as implied by these structured priors; therefore MCMC methods for parameter estimation are necessary. Prior on latent component structure were introduced for univariate AR models in Huerta and West (1999). In this sense, the models proposed in this paper are an extension to the multivariate framework.

2 Multivariate time series decompositions

In this section, we describe general time series decomposition results for a class of multivariate time series processes. We discuss such results in detail for the particular case of what we call *diagonal vector autoregressions* or DVARs. Similar to the univariate case, the decomposition results summarized below provide a natural framework for the structured prior specification that is developed in section 3. Further details and applications related to decompositions for univariate autoregressions and time-varying autoregressions can be found in West (1997), Huerta and West (1999), West *et al.* (1999) and Prado and Huerta (2002). Here, we revisit the developments on multivariate time series decompositions presented in Prado (1998) and include extensions that handle a more general model case.

Consider an m-dimensional time series process $\mathbf{y}_t = (y_{1,t}, \ldots, y_{m,t})'$ that is modeled with a multivariate dynamic linear model or MDLM (West and Harrison, 1997)

$$\mathbf{y}_t = \mathbf{x}_t + \boldsymbol{\nu}_t, \quad \mathbf{x}_t = \mathbf{F}'\boldsymbol{\theta}_t, \quad \boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \tag{1}$$

where $\mathbf{x}_t$ is the underlying m-dimensional signal, $\boldsymbol{\nu}_t$ is an m-dimensional vector of observation errors, $\mathbf{F}'$ is an $m \times d$ matrix of constants, $\boldsymbol{\theta}_t$ is the d-dimensional state vector, $\mathbf{G}_t$ is the $d \times d$ state evolution matrix and $\boldsymbol{\omega}_t$ is a d-vector of state innovations. The noise terms $\boldsymbol{\nu}_t$ and $\boldsymbol{\omega}_t$ are zero mean innovations, assumed independent and mutually independent with variance-covariance matrices $\mathbf{V}_t$ and $\mathbf{W}_t$ respectively.

A scalar DLM can be written for each of the univariate components of $\mathbf{x}_t$, namely

$$\mathcal{M}_i: \quad \begin{array}{rcl} x_{i,t} & = & \mathbf{F}_i'\boldsymbol{\theta}_t \\ \boldsymbol{\theta}_t & = & \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \end{array} \tag{2}$$

with $\mathbf{F}_i$ the i-th column of the matrix $\mathbf{F}$. Each scalar component $x_{i,t}$ of the m-dimensional signal vector can be broken into latent processes using the decomposition results for univariate time series presented in West *et al.* (1999). Assume that the system evolution matrix $\mathbf{G}_t$ is diagonalizable, i.e. that there exist a diagonal matrix $\mathbf{A}_t$, and a matrix $\mathbf{B}_t$ such that $\mathbf{G}_t = \mathbf{B}_t\mathbf{A}_t\mathbf{B}_t^{-1}$. A useful way to characterize a diagonalizable matrix is by the multiplicities of its eigenvalues. If $\mathbf{G}_t$ has $d^* \leq d$ distinct eigenvalues, $\lambda_{t,1}, \ldots, \lambda_{t,d^*}$ with algebraic multiplicities $m_{a,1}, \ldots, m_{a,d^*}$ respectively, then $\mathbf{G}_t$ is diagonalizable if and only if $m_{g,i} = m_{a,i}$ for all $i = 1, \ldots, d^*$, with $m_{g,i}$ the geometric multiplicity of the eigenvalue $\lambda_{t,i}$. That is, $\mathbf{G}_t$ is diagonalizable if and only if the algebraic multiplicity of each eigenvalue equals its geometric multiplicity. In particular, if $\mathbf{G}_t$ has exactly d distinct eigenvalues, then $\mathbf{G}_t$ is diagonalizable. Note we are assuming that the number of distinct eigenvalues d^*, the number of real and complex eigenvalues and their multiplicities remain fixed over time. In other words, we assume that there are exactly c^* pairs of distinct complex eigenvalues $r_{t,j}\exp(\pm i\omega_{t,j})$ for $j = 1, \ldots, c^*$, and $r^* = d^* - 2c^*$ distinct real eigenvalues for $j = 2c^* + 1, \ldots, d^*$ at each time t. Then, $\mathbf{G}_t = \mathbf{B}_t\mathbf{A}_t\mathbf{B}_t^{-1}$ with $\mathbf{A}_t$ the $d \times d$ diagonal matrix of eigenvalues, in arbitrary but fixed order, and $\mathbf{B}_t$ a corresponding matrix of eigenvectors. For each t and each model $\mathcal{M}_i$ define the matrices $\mathbf{H}_{i,t} = diag(\mathbf{B}_t'\mathbf{F}_i)\mathbf{B}_t^{-1}$ for $i = 1, \ldots, m$, and reparameterize $\mathcal{M}_i$ via $\boldsymbol{\gamma}_{i,t} = \mathbf{H}_{i,t}\boldsymbol{\theta}_t$ and $\boldsymbol{\delta}_{i,t} = \mathbf{H}_{i,t}\boldsymbol{\omega}_t$. Then, rewriting (2) in terms of the new state and innovation vectors, we have

$$\begin{array}{rcl} x_{i,t} & = & \mathbf{1}'\boldsymbol{\gamma}_{i,t} \\ \boldsymbol{\gamma}_{i,t} & = & \mathbf{A}_t\mathbf{K}_{i,t}\boldsymbol{\gamma}_{i,t-1} + \boldsymbol{\delta}_{i,t}, \end{array} \tag{3}$$

where $\mathbf{1}' = (1, \ldots, 1)$ and $\mathbf{K}_{i,t} = \mathbf{H}_{i,t}\mathbf{H}_{i,t-1}^{-1}$. Therefore $x_{i,t}$ can be expressed as a sum of d^* components

$$x_{i,t} = \sum_{j=1}^{c^*} z_{i,t,j} + \sum_{j=2c^*+1}^{d^*} y_{i,t,j}, \tag{4}$$

where $z_{i,t,j}$ are real-valued processes related to the pairs of complex eigenvalues given by $r_{t,j}\exp(\pm i\omega_{t,j})$ for $j = 1, \ldots, c^*$, and $y_{i,t,j}$ are real processes related to the real eigenvalues $r_{t,j}$ for $j = 2c^* + 1, \ldots, d^*$.

2.1 Decomposition of the scalar components in a VAR(p)

Consider the particular case of an m-dimensional time series process $\mathbf{x}_t = (x_{1,t}, \ldots, x_{m,t})'$ that follows a VAR(p)

$$\mathbf{x}_t = \mathbf{\Phi}_1 \mathbf{x}_{t-1} + \mathbf{\Phi}_2 \mathbf{x}_{t-1} + \ldots + \mathbf{\Phi}_p \mathbf{x}_{t-p} + \boldsymbol{\epsilon}_t, \tag{5}$$

where $\mathbf{\Phi}_j$ are the $m \times m$ matrices of AR coefficients and $\boldsymbol{\epsilon}_t$ is the m-dimensional zero mean innovation vector at time t, with covariance matrix Σ. The VAR(p) process in (5) is *stable* (see for instance Lütkepohl, 1993), if the polynomial $\mathbf{\Phi}(u) = det(\mathbf{I}_m - \mathbf{\Phi}_1 u - \ldots - \mathbf{\Phi}_p u^p)$, with $\mathbf{I}_m$ the $m \times m$ identity matrix, has no roots within or on the complex unit circle. If a VAR(p) process is stable then it is stationary.

Any m-dimensional VAR(p) process can be written in the MDLM form (1), with $d = mp$, $\boldsymbol{\nu}_t = 0$, and the $m \times (mp)$ matrix of constants $\mathbf{F}'$ and the (mp)-dimensional state and the state innovation vectors $\boldsymbol{\theta}_t$ and $\boldsymbol{\omega}_t$ described by

$$\mathbf{F}' = \begin{pmatrix} \mathbf{e}_1' & 0 & \ldots & 0 \\ \mathbf{e}_2' & 0 & \ldots & 0 \\ \vdots & & & \vdots \\ \mathbf{e}_m' & 0 & \ldots & 0 \end{pmatrix}; \quad \boldsymbol{\theta}_t = \begin{pmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \\ \vdots \\ \mathbf{x}_{t-p+1} \end{pmatrix}; \quad \boldsymbol{\omega}_t = \begin{pmatrix} \boldsymbol{\epsilon}_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{6}$$

where each $\mathbf{e}_j$ is an m-dimensional vector whose j-th element is equal to unity and all the other elements are zeros. Finally, the $(mp) \times (mp)$ state evolution matrix $\mathbf{G}$ is given by

$$\mathbf{G} = \begin{pmatrix} \mathbf{\Phi}_1 & \mathbf{\Phi}_2 & \ldots & \mathbf{\Phi}_{p-1} & \mathbf{\Phi}_p \\ \mathbf{I}_m & \mathbf{0}_m & \ldots & \mathbf{0}_m & \mathbf{0}_m \\ \vdots & \ddots & & & \vdots \\ \mathbf{0}_m & \mathbf{0}_m & \ldots & \mathbf{I}_m & \mathbf{0}_m \end{pmatrix}, \tag{7}$$

with $\mathbf{0}_m$ the $m \times m$ dimensional matrix of zeros. The eigenvalues of $\mathbf{G}$ satisfy the equation

$$det(\mathbf{I}_m \lambda^p - \mathbf{\Phi}_1 \lambda^{p-1} - \mathbf{\Phi}_2 \lambda^{p-2} - \ldots - \mathbf{\Phi}_p) = 0,$$

i.e. they are the reciprocal roots of the polynomial $\mathbf{\Phi}(u)$. Therefore, $\mathbf{x}_t$ is stable if the eigenvalues of $\mathbf{G}$ have modulus less than one. Assume that $\mathbf{G}$ has $d^* \leq mp$ distinct eigenvalues with c^* pairs of distinct complex eigenvalues $r_j \exp(\pm i\omega_j)$ for $j = 1, \ldots, c^*$, and $r^* = d^* - 2c^*$ real eigenvalues r_j for $j = 2c^* + 1, \ldots, d^*$. If $\mathbf{G}$ is diagonalizable, then, using the representations (2) and (3), and the fact that $\mathbf{K}_{i,t} = \mathbf{I}$ for all i, j we have

$$x_{i,t} = \sum_{j=1}^{c^*} z_{i,t,j} + \sum_{j=2c^*+1}^{d^*} y_{i,t,j}, \tag{8}$$

where each $z_{i,t,j}$ is quasi-periodic following an ARMA(2,1) model with characteristic modulus r_j and frequency ω_j for all $i = 1, \ldots, m$. Then, the moduli and frequencies that characterize the processes $z_{i,t,j}$ for a fixed j, are the same across the m univariate series that define the VAR process. Similarly, $y_{i,t,j}$ is an AR(1) process whose AR coefficient is the real eigenvalue r_j for all $i = 1, \ldots, m$.

Example. *Vector autoregressions with diagonal matrices of coefficients or DVAR(p).* Suppose that we have an m-dimensional VAR(p) process with $\mathbf{\Phi}_j = diag(\phi_{1,j}, \ldots, \phi_{m,j})$ for $j = 1, \ldots, p$. Then, the characteristic polynomial of the process is given by

$$\mathbf{\Phi}(u) \;=\; \prod_{i=1}^{m}(1 - \phi_{i,1}u - \phi_{i,2}u^2 - \ldots - \phi_{i,p}u^p) = \prod_{i=1}^{m}\mathbf{\Phi}^i(u),$$

i.e. $\mathbf{\Phi}(u)$ is the product of the characteristic polynomials associated to each of the m series. Let $\alpha_1^1, \ldots, \alpha_p^1, \ldots, \alpha_1^m, \ldots, \alpha_p^m$ be the reciprocal roots of the characteristic polynomials $\mathbf{\Phi}^1(u), \ldots, \mathbf{\Phi}^m(u)$, respectively, with $\alpha_j^i \neq 0$ for all i, j. Assume that for a fixed series i, the reciprocal roots α_j^i are all distinct, but common roots across series are allowed, that is $\alpha_j^i = \alpha_l^k$ for some i, k such that $i \neq k$ and some j, l. If there are c^* distinct complex pairs of reciprocal roots, denoted by $r_j \exp(\pm i\omega_j)$ for $j = 1, \ldots, c^*$, r^* pairs of distinct real roots r_j, for $j = 2c^* + 1, \ldots, d^*$ with $2c^* + r^* = d^* \leq mp$, and $\mathbf{G}$ is diagonalizable, then the decomposition (8) holds. It is easy to see that the state evolution matrix $\mathbf{G}$ in this case is diagonalizable by showing that, for any eigenvalue $\lambda \neq 0$ of $\mathbf{G}$, its algebraic multiplicity $m_{a,\lambda}$ equals its geometric multiplicity $m_{g,\lambda}$, with $m_{g,\lambda}$ the dimension of the characteristic subspace of λ, $\{\mathbf{x} : (\mathbf{G} - \lambda\mathbf{I}_{mp})\mathbf{x} = \mathbf{0}_{mp}\}$.

3 The prior structure

We extend the priors on autoregressive root structure developed in Huerta and West (1999), to the context of vector autoregressions with diagonal matrices of coefficients or DVARs. Assume that we have an m-dimensional series. We begin by specifying fixed upper bounds C_i and R_i on the number of complex root pairs and real roots of series i, for $i = 1, \ldots, m$. Conditional on these upper bounds, we assume a prior structure on the component roots α_j^i for $j = 1, \ldots, 2C_i + R_i$, that distinguishes between real and complex cases. Let us introduce some notation that will be useful to define the prior structure.

- r_j^i and $\lambda_j^i = 2\pi/\omega_j^i$ are the modulus and the wavelength or period of the j-th component root of series i;

- $\pi_{r^i,(x|x\neq x_1,\ldots,x_K)}$ denotes the prior probability that a given modulus related to the series i, takes a value of x conditional on x being different from the values $x_1, \ldots, x_K$. Similarly, $\pi_{\lambda^i,(x|x\neq x_1,\ldots,x_K)}$ denotes the prior probability that a given period related to the series i takes a value of x conditional on x being different from the values $x_1, \ldots, x_K$;

- $\mathbf{r}^i_{1:j} = \{r^i_1, \ldots, r^i_j\}$; $\boldsymbol{\lambda}^i_{1:j} = \{\lambda^i_1, \ldots, \lambda^i_j\}$; $\boldsymbol{\alpha}^i_{1:j} = (\mathbf{r}, \boldsymbol{\lambda})^i_{1:j} = \{(r^i_1, \lambda^i_1), \ldots, (r^i_j, \lambda^i_j)\}$;
- $I_y(z)$ is the indicator function, i.e.,$I_y(z) = 1$ if $z = y$ and 0 otherwise;
- $U(x|a, b)$ denotes a Uniform distribution on x over the interval (a, b) and $Bn(x|n, p)$ denotes a Binomial distribution on x with parameters n and p.

Then, we assume the following prior structure on the component roots of the m series.

(a) **Real roots.**

Let $p(\mathbf{r}^1_{1:R_1}, \ldots, \mathbf{r}^m_{1:R_m})$ be the prior on the real roots for the m series. Then, the following structure is proposed

$$p(\mathbf{r}^1_{1:R_1}, \ldots, \mathbf{r}^m_{1:R_m}) = p(\mathbf{r}^1_{1:R_1}) \times p(\mathbf{r}^2_{1:R_2}|\mathbf{r}^1_{1:R_1}) \times \cdots \times p(\mathbf{r}^m_{1:R_m}|\mathbf{r}^1_{1:R_1}, \ldots, \mathbf{r}^{m-1}_{1:R_{m-1}}),$$

with $p(\mathbf{r}^1_{1:R_1}) = \prod_{j=1}^{R_1} p(r^1_j)$, i.e., conditional on the upper bound R_1, we assume independent priors on the component roots of the first series. Now, for each series k, with $k = 2, \ldots, m$, we assume a conditionally independent structure on the component roots of such series, that is,

$$p(\mathbf{r}^k_{1:R_k}|\mathbf{r}^1_{1:R_1}, \ldots, \mathbf{r}^{k-1}_{1:R_{k-1}}) = \prod_{j=1}^{R_k} p(r^k_j|\mathbf{r}^1_{1:R_1}, \ldots, \mathbf{r}^{k-1}_{1:R_{k-1}}, \mathbf{r}^k_{1:(j-1)}).$$

Therefore, begin by specifying a fixed upper bound R_1. Then, following Huerta and West (1999), the real root r^1_j for $j = 1, \ldots, R_1$ has a prior with support $|r^1_j| < 1$ and density

$$r^1_j \sim \pi_{r^1,-1} I_{(-1)}(r^1_j) + \pi_{r^1,0} I_0(r^1_j) + \pi_{r^1,1} I_1(r^1_j) + (1 - \pi_{r^1,-1} - \pi_{r^1,0} - \pi_{r^1,1}) g_r(r^1_j),$$

where $g_r(\cdot)$ is a continuous density over $(-1, 1)$, and $\pi_{r^1,\cdot}$ are prior probabilities at $r^1_j = -1, 0$ and 1. The prior point mass at $r^1_j = 0$ allows to reduce the number of real roots of the first series below the specified upper bound R_1, while the point masses at $r^1_j = \pm 1$ allow for roots on the stationary boundary. The continuous part of the mixture prior specifies the conditional prior on the full stationary region $|r^1_j| < 1$. For instance, the reference prior is the uniform, $g_r(\cdot) = U(\cdot| - 1, 1)$. Assuming that we have fixed values of $\pi_{r^1,\cdot}$, we have an implied Binomial prior $Bn(R_1, 1 - \pi_{r^1,0})$ on the number of non-zero roots.

Now, the priors on the autoregressive structure for the real components of the second series are conditional on the real root components of the first series. Then, given a specified upper bound R_2, r^2_j for $j = 1, \ldots, R_2$ has prior over support $|r^2_j| < 1$

$$\begin{aligned} r^2_j|\mathbf{r}^1_{1:R_1}, \mathbf{r}^2_{1:(j-1)} \quad \sim \quad & \sum_{k=1}^{R^1} \pi_{r^2,(r^1_k|r^1_k \neq 0,\pm 1,\mathbf{r}^2_{1:(j-1)})} I_{r^1_k}(r^2_j) + \\ & \pi_{r^2,-1} I_{(-1)}(r^2_j) + \pi_{r^2,0} I_0(r^2_j) + \pi_{r^2,1} I_1(r^2_j) + \\ & \left(1 - \sum_{k=1}^{R^1} \pi_{r^2,(r^1_k|r^1_k \neq 0 \pm 1,\mathbf{r}^2_{1:(j-1)})} - \sum_{q=-1,0,1} \pi_{r^2,q}\right) g_r(r^2_j). \end{aligned}$$

Here the $\pi_{r^2,\cdot}$ are prior probabilities at $r_j = r_k^1$ for $k = 1, \ldots, R_1$ provided that r_k^1 is neither 0, 1 or -1, nor a root previously considered for the second series. This structure permits to have non-zero prior probabilities at roots previously sampled for the first series, allowing for repeated roots across series. Again, $g_r(\cdot)$ is a continuous density over $(-1, 1)$. Following the previous structure for $i = 3, \ldots, m$ and fixed upper bounds R_i on the maximum number of real roots, r_j^i has prior over the support $|r_j^i| < 1$ with density

$$\begin{aligned} r_j^i | \mathbf{r}_{1:R_1}^1, \ldots, \mathbf{r}_{1:R_{i-1}}^{i-1}, \mathbf{r}_{1:(j-1)}^i \quad \sim \quad & \sum_{l=1}^{i-1} \sum_{k=1}^{R^l} \pi_{r^i,(r_k^l | r_k^l \neq 0 \pm 1, \mathbf{r}_{1:(j-1)}^i)} I_{r_k^l}(r_j^i) + \\ & \pi_{r^i,-1} I_{(-1)}(r_j^i) + \pi_{r^i,0} I_0(r_j^i) + \pi_{r^i,1} I_1(r_j^i) + \\ & \left(1 - \sum_{l=1}^{i-1} \sum_{k=1}^{R^l} \pi_{r^i,(r_k^l | r_k^l \neq 0, \pm 1, \mathbf{r}_{1:(j-1)}^i)} - \sum_{q=-1,0,1} \pi_{r^i,q}\right) g_r(r_j^i). \end{aligned}$$

(b) **Complex roots.**

The prior for the complex roots has a similar structure to the prior for the real roots. Again, we start specifying upper bounds C_i on the number of complex roots pairs of each series and take the defining parameters $\alpha_j^i = (r_j^i, \lambda_j^i)$ where r_j^i is the modulus and $\lambda_j^i = 2\pi/\omega_j^i$ is the wavelength of the quasi-periodic component $z_{t,i,j}$ corresponding to this root pair. Let $p(\boldsymbol{\alpha}_{1:C_1}^1, \ldots, \boldsymbol{\alpha}_{1:C_m}^m)$ be the prior on the complex root pairs for the m series. Then, the following structure is proposed

$$p(\boldsymbol{\alpha}_{1:C_1}^1, \ldots, \boldsymbol{\alpha}_{1:C_m}^m) \quad = \quad p(\boldsymbol{\alpha}_{1:C_1}^1) \times p(\boldsymbol{\alpha}_{1:C_2}^2 | \boldsymbol{\alpha}_{1:C_1}^1) \times \cdots \times p(\boldsymbol{\alpha}_{1:C_m}^m | \boldsymbol{\alpha}_{1:C_1}^1 \ldots, \boldsymbol{\alpha}_{1:C_{m-1}}^{m-1}),$$

with $p(\boldsymbol{\alpha}_{1:C_1}^1) = \prod_{j=1}^{C_1} p(r_j^1) p(\lambda_j^1)$. Thus, conditional on the upper bound C_1, we assume independent priors on the component roots of the first series such that the moduli r_j^1 and the wavelengths λ_j^1 are independent for all j. The prior is specified over the support $0 \leq r_j^1 \leq 1$ and $2 < \lambda_j^1 < \lambda_u$ for a given upper bound λ_u on the wavelengths. For each modulus r_j^1, $j = 1, \ldots, C_1$, we assume

$$r_j^1 \sim \pi_{c^1,0} I_0(r_j^1) + \pi_{c^1,1} I_1(r_j^1) + (1 - \pi_{c^1,0} - \pi_{c^1,1}) g_c(r_j^1).$$

The first component corresponds to a zero root, allowing the number of complex components to be less than C_1. Again, the second component allows for roots on the stationary boundary. The component $g_c(r_j^1)$ is a continuous distribution on $0 < r_j^1 < 1$, the conditional prior on the stationary region. The prior on each $\alpha_j^1 = (r_j^1, \lambda_j^1)$ is completed by specifying a marginal density $h_1(\lambda_j^1)$ for the wavelength over the support $(2, \lambda_u)$. Different choices of $g_c(r_j^1)$ and $h_1(\lambda_j^1)$ can be considered, including Uniform priors and marginals for λ_j^1 based on Uniform priors for the corresponding frequency or "component reference priors" (Huerta and West, 1999). The priors on the autoregressive structure for the complex pairs of the second series, $\alpha_j^2 = (r_j^2, \lambda_j^2)$, are conditional on the complex root components of the first series and on the complex roots previously

sampled for the second series, i.e., given an upper bound C_2, r_j^2 has prior over the support $0 < r_j^2 < 1$ with,

$$\begin{aligned} r_j^2|\mathbf{r}_{1:C_1}^1, \mathbf{r}_{1:(j-1)}^2 &\sim \sum_{k=1}^{C_1} \pi_{c^2,(r_k^1|r_k^1\neq 0,1,\mathbf{r}_{1:(j-1)}^2)} I_{r_k^1}(r_j^2) + \pi_{c^2,0}(r_j^2) + \pi_{c^2,1}(r_j^2) \\ &\quad + \left(1 - \pi_{c^2,0} - \pi_{c^2,1} - \sum_{k=1}^{C_1} \pi_{c^2,(r_k^1|r_k^1\neq 0,1,\mathbf{r}_{1:(j-1)}^2)}\right) g_c(r_j^2), \end{aligned}$$

and the period λ_j^2 has prior over support $(2, \lambda_u)$ with,

$$\lambda_j^2|\boldsymbol{\alpha}_{1:C_1}^1, \boldsymbol{\alpha}_{1:(j-1)}^2 \sim \sum_{k=1}^{C_1} I_{r_k^1}(r_j^2) I_{\lambda_k^1}(\lambda_j^2) + \left(1 - \sum_{k=1}^{C_1} I_{r_k^1}(r_j^2) I_{\lambda_k^1}(\lambda_j^2)\right) h(\lambda_j^2).$$

Then, for each series k, with $k = 2, \ldots, m$ we assume the following conditionally independent prior structure specified over support $0 < r_j^i < 1$ and $2 < \lambda_j^i < \lambda_u$,

$$\begin{aligned} r_j^i|\mathbf{r}_{1:C_1}^1, \ldots, \mathbf{r}_{1:C_{i-1}}^{i-1}, \mathbf{r}_{1:(j-1)}^i &\sim \sum_{l=1}^{i-1}\sum_{k=1}^{C_l} \pi_{c^i,(r_k^l|r_k^l\neq 0,1,\mathbf{r}_{1:(j-1)}^i)} I_{r_k^l}(r_j^i) + \pi_{c^i,0} I_0(r_j^i) + \pi_{c^i,1} I_1(r_j^i) \\ &\quad + \left(1 - \pi_{c^i,0} - \pi_{c^i,1} - \sum_{l=1}^{i-1}\sum_{k=1}^{C_l} \pi_{c^i,(r_k^l|r_k^l\neq 0,1,\mathbf{r}_{1:(j-1)}^i)}\right) g_c(r_j^i), \end{aligned}$$

and for the periods of the complex roots,

$$\lambda_j^i|\boldsymbol{\alpha}_{1:C_1}^i, \ldots, \boldsymbol{\alpha}_{1:C_{i-1}}^{i-1}, \boldsymbol{\alpha}_{1:(j-1)}^i \sim \sum_{l=1}^{i-1}\sum_{k=1}^{C_l} I_{r_k^l}(r_j^i) I_{\lambda_k^l}(\lambda_j^i) + \left[1 - \sum_{l=1}^{i-1}\sum_{k=1}^{C_l} I_{r_k^l}(r_j^i) I_{\lambda_k^l}(\lambda_j^i)\right] h(\lambda_j^i).$$

3.1 Some aspects of implied prior structure

The priors specified on the roots structure induce priors on the numbers of complex and real roots associated with each series, and so on model order up to the specified maximum. These priors also induce priors, of complicated mathematical forms, on the standard linear autoregressive parameters $\phi_{i,k}$, for $i = 1, \ldots, m$ and $k = 1, \ldots, p$. Consider for instance a $VAR_2(4)$ model with exactly two real components in each series $R_1 = R_2 = 2$, and one quasi-periodic component in each series, $C_1 = C_2 = 1$, taking $\pi_{r^1,0} = \pi_{r^2,0} = \pi_{c^1,0} = \pi_{c^2,0} = 0$, $\pi_{r^1,-1} = \pi_{r^2,-1} = 0$ and $\pi_{r^1,1} = \pi_{r^2,1} = \pi_{c^1,1} = \pi_{c^2,1} = 0$. In addition, we take $g_r(r_j^i)$, $g_c(r_j^i)$ and $h(\lambda_j^i)$ as Uniform distributions. A discrete Uniform distribution is set on the weights $\pi_{r^i,\cdot}$ and $\pi_{c^i,\cdot}$ that are not equal to zero for each i. We explore the implied prior on the eight AR coefficients $\boldsymbol{\phi} = (\phi_{1,1}, \ldots, \phi_{1,4}, \phi_{2,1}, \ldots, \phi_{2,4})$ via simulation: given a random draw from the prior, we can trivially compute the corresponding value of $\boldsymbol{\phi}$. Figure 1 displays two-dimensional margins of a sample of 10,000 draws from the prior. The two-dimensional margins of the

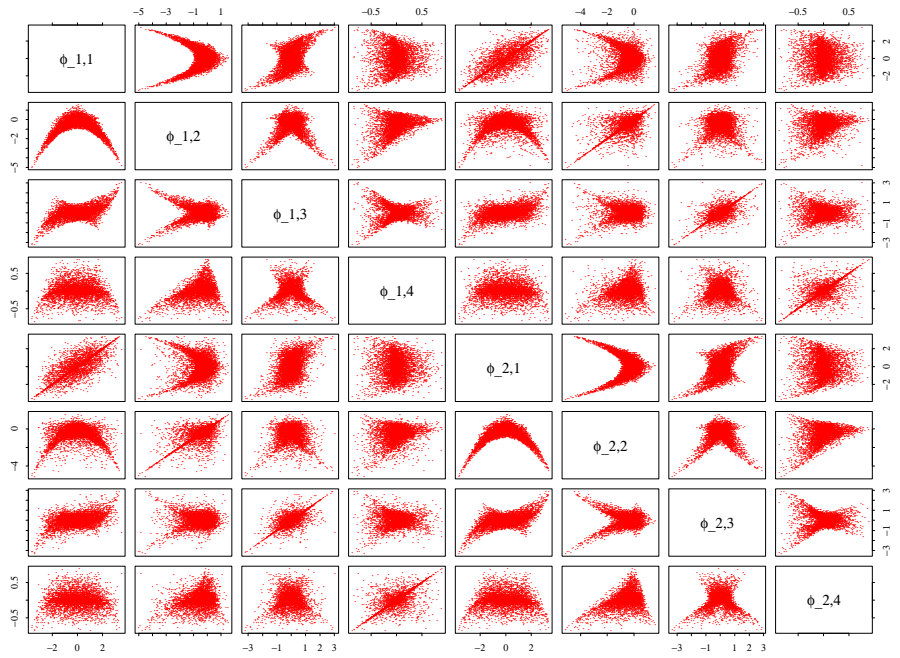

Figure 1: *Samples from the prior for ϕ in a DVAR$_2$(4) model with $R_1 = R_2 = 2$ and $C_1 = C_2 = 1$.*

AR coefficients $\boldsymbol{\phi}_1 = (\phi_{1,1}, \ldots, \phi_{1,4})$ and $\boldsymbol{\phi}_2 = (\phi_{2,1}, \ldots, \phi_{2,4})$ (see four by four diagonal picture blocks in the Figure) show similar displays to the ones of two-dimensional margins obtained via the implied structured priors on the four coefficients using a standard univariate AR(4) model (see Huerta and West 1999). The two four by four off diagonal blocks in the Figure show the correlation structure between the AR coefficients $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$. By construction, the prior for $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ is constrained to the stationary region and so the shapes in Figure 1 are contained in this region. Note that, as in the univariate case, the induced prior on $\boldsymbol{\phi}$ is naturally not uniform.

4 Posterior structure in DVAR models

Under the prior structure just described, posterior and predictive calculations are available via Markov chain Monte Carlo (MCMC) simulation methods (e.g., Gamerman, 1997). The structure of relevant conditional posterior distributions is briefly outlined here, details will be reported elsewhere.

Assume we have m series and let $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, with $\mathbf{x}_t = (x_{1,t}, \ldots, x_{m,t})'$ be the observed multivariate time series and, given the maximum model order $p = \max\{p_i, i = 1, \ldots, m\}$, write $\mathbf{X}_0 = \{\mathbf{x}_0, \mathbf{x}_{-1}, \ldots, \mathbf{x}_{-(p-1)}\}$ for the latent initial values. Let Σ be the $m \times m$ variance-covariance matrix. The model parameters are denoted

by $\boldsymbol{\alpha} = \{\alpha_1^1, \ldots, \alpha_{p_1}^1, \ldots, \alpha_1^m, \ldots, \alpha_{p_m}^m\}$. Assuming Σ and $\mathbf{X}_0$ known, the posterior inferences are based on summarizing the full posterior $p(\boldsymbol{\alpha}, |\mathbf{X}_0, \mathbf{X}, \Sigma)$. For any subset $\boldsymbol{\xi}$ of elements of $\boldsymbol{\alpha}$, let $\boldsymbol{\alpha}\backslash\boldsymbol{\xi}$ denote the complementary elements, that is, $\boldsymbol{\alpha}$ with $\boldsymbol{\xi}$ removed. Our MCMC method is based on a standard Gibbs sampling format, specifically

- for each $i = 1, \ldots, m$,
  - for each $j = 2C_i + 1, \ldots, 2C_i + R_i$, sample the real roots individually from $p(r_j^i|\boldsymbol{\alpha}\backslash r_j^i, \mathbf{X}, \mathbf{X}_0, \Sigma)$;
  - for each $j = 1, \ldots, C_i$, sample the complex roots individually from the full conditional, $p(\alpha_j^i|\boldsymbol{\alpha}\backslash\alpha_j^i, \mathbf{X}, \mathbf{X}_0, \Sigma)$.

Each of these distributions is now briefly described.

(a) *Conditional distributions for real roots.* Consider any real root $\alpha_j^i = r_j^i$, for some series i and some j between $2C_i + 1$ and p_i. Given $\boldsymbol{\alpha}\backslash r_j^i$, $\mathbf{X}$, $\mathbf{X}_0$ and the DVAR model, the likelihood function for r_j^i provides a normal kernel in r_j^i. Under this mixture prior, this leads to the mixture posterior

$$\sum_{l=1}^{i-1}\sum_{k=1}^{R^l} p_{j,r_k^l}^i I_{r_k^l}(r_j^i) + \sum_{q=-1,0,1} p_{j,q} I_q(r_j^i) + (1 - \sum_{q=-1,0,1} p_{j,q} - \sum_{l=1}^{i-1}\sum_{k=1}^{R^l} p_{j,r_k^l}^i) N_t(r_j^i|m_j^i, M_j^i)$$

where $N_t(\cdot|m, M)$ denotes the density of the normal $N(\cdot|m, M)$ truncated to $(-1, 1)$, and the values (m_j^i, M_j^i) and point masses are trivially computed. This mixture posterior is easily sampled with direct simulation of the truncated normal by c.d.f. inversion.

(b) *Conditional for complex roots.* For each i, index $j = 1, \ldots, C_i$ identifies the pair of complex conjugate roots $(\alpha_{2j-1}^i, \alpha_{2j}^i)$ with parameters (r_j^i, λ_j^i). Let A_j^i be the index set of all other roots, $\boldsymbol{\alpha}\backslash(r_j^i, \lambda_j^i)$. Then, given $\boldsymbol{\alpha}\backslash(r_j^i, \lambda_j^i)$ and $\mathbf{X}$ we can directly compute the filtered time series, $z_{t,l} = \prod_{k\in A_j^i}(1 - \alpha_k^l B)x_{t,l}$ if $l = i$ and $z_{t,l} = \prod_{k=1}^{p_l}(1 - \alpha_j^i B)x_{t,l}$. Now, the likelihood on $\phi_{j,1}^i = 2r_j^i \cos(2\pi/\lambda_j^i)$ and $\phi_{j,2}^i = -(r_j^i)^2$ provides a bivariate normal kernel with a mean vector and a variance-covariance matrix that are functions of the filtered time series $z_{t,1}, \ldots, z_{t,m}$. However, sampling from the resulting conditional posterior directly is difficult and because of this, following Huerta and West (1999), we use a reversible jump Markov chain Monte Carlo step.

The structure of the MCMC algorithm in this case is very similar to the structure of the MCMC algorithm developed in Huerta and West (1999) for the univariate case. However, the number of computations increases considerably when the number of series and/or the model order are large.

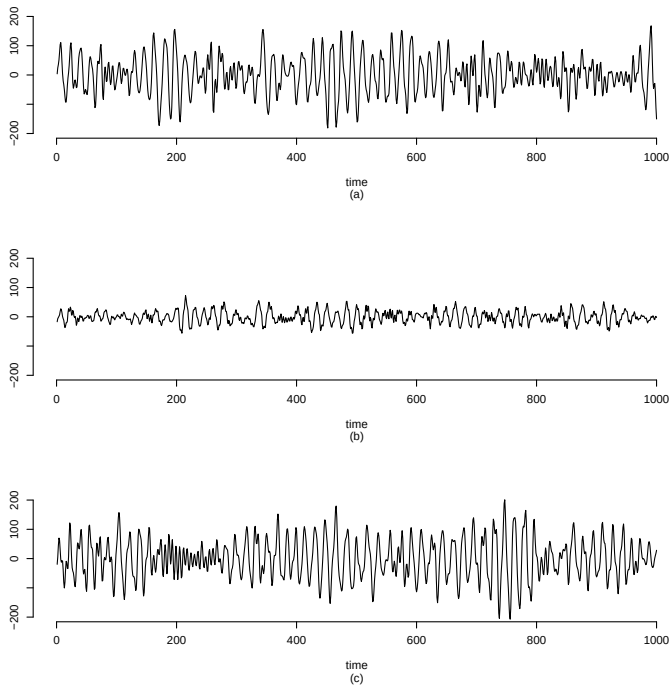

Figure 2: *Simulated series. Graph (a) corresponds to a process with two pairs of roots with moduli* $r_1^1 = 0.98$, $r_2^1 = 0.97$ *and wavelengths* $\lambda_1^1 = 16.75$ *and* $\lambda_2^1 = 6.28$, *respectively. Graph (b) corresponds to a process with roots having moduli and wavelengths of* $r_1^2 = 0.98$, $r_2^2 = 0.97$, $\lambda_1^2 = 16.75$ *and* $\lambda_2^2 = 4.0$. *The series in graph (c) was simulated using the same root structure used to generate the series displayed in (a).*

5 A synthetic data example: a DVAR$_3(4)$ with two pairs of complex roots

In this section, we analyze simulated data from a three-dimensional diagonal vector AR process with two pairs of complex roots and variance-covariance matrix $\Sigma = \mathbf{I}_3$. Figure (2) displays 1,000 data points for each of the three simulated series. The first series corresponds to an autoregressive process with two pairs of complex roots with moduli and wavelengths of $r_1^1 = 0.98$, $r_2^1 = 0.97$ and $\lambda_1^1 = 16.75$, $\lambda_2^1 = 6.28$, respectively. The third series has the same root structure as the first series. The second series has one common pair of roots with the first and the third series, namely $r_1^2 = 0.98$ and $\lambda_1^2 = 16.75$, and another pair characterized by $r_2^2 = 0.97$ and $\lambda_2^2 = 4.0$. We assume a prior structure with a maximum of two pairs of complex roots $C_i = 2$, for each series $i = 1, 2, 3$ and no real roots, i.e. $R_i = 0$ for all i. The prior masses for the roots on the stationary boundary were set to zero and a discrete Uniform prior was used for the prior masses of roots in the stationary region. In addition, we take a discrete Uniform

Series	Marginal posterior probabilities	
$i=1$	$Pr(r_1^1=0\|\mathbf{X})=0$	$\boldsymbol{Pr(r_1^1\sim C\|\mathbf{X})=1.000}$
	$Pr(r_2^1=0\|\mathbf{X})=0$	$\boldsymbol{Pr(r_2^1\sim C\|\mathbf{X})=1.000}$
$i=2$	$Pr(r_1^2=r_1^1\|\mathbf{X})=0.321$	$Pr(r_1^2=r_2^1\|\mathbf{X})=0.074$
	$Pr(r_1^2=0\|\mathbf{X})=0.015$	$\boldsymbol{Pr(r_1^2\sim C\|\mathbf{X})=0.590}$
	$Pr(r_2^2=r_1^1\|\mathbf{X})=0.000$	$Pr(r_2^2=r_2^1\|\mathbf{X})=0.069$
	$Pr(r_2^2=0\|\mathbf{X})=0.000$	$\boldsymbol{Pr(r_2^2\sim C\|\mathbf{X})=0.931}$
$i=3$	$\boldsymbol{Pr(r_1^3=r_1^1\|\mathbf{X})=0.610}$	$Pr(r_1^3=r_2^1\|\mathbf{X})=0.021$
	$Pr(r_1^3=r_1^2\|\mathbf{X})=0.083$	$Pr(r_1^3=r_2^2\|\mathbf{X})=0.002$
	$Pr(r_1^3=0\|\mathbf{X})=0.062$	$Pr(r_1^3\sim C\|\mathbf{X})=0.222$
	$Pr(r_2^3=r_1^1\|\mathbf{X})=0.010$	$\boldsymbol{Pr(r_2^3=r_2^1\|\mathbf{X})=0.474}$
	$Pr(r_2^3=r_1^2\|\mathbf{X})=0.080$	$Pr(r_2^3=r_2^3\|\mathbf{X})=0.238$
	$Pr(r_2^3=0\|\mathbf{X})=0.000$	$Pr(r_2^3\sim C\|\mathbf{X})=0.198$

Table 1: *Marginal posterior distributions*

prior on $\pi_{c^i,\cdot}$ for each series, while $g_c(\cdot)$ and $h(\cdot)$ are taken as component reference priors (see Huerta and West, 1999). The posterior summaries presented here are based on a sample of 1,000 draws taken from 10,000 iterations of the Gibbs sampler, described in the previous section, after a burn-in period of 10,000 iterations for MCMC convergence.

For this particular example, the number of possible models is 504, which is a large number, considering we have a small number of series and a small model order. The number of possible models increases enormously when the number of series and/or AR model orders for each series increases, so exploring the posterior distributions obtained under the proposed class of priors is not trivial. In this case, we can explore exhaustively the marginal posterior probabilities for all the possible roots. Table 1 displays the results of the marginal posterior probabilities of the zero roots $Pr(r_j^i = 0|\mathbf{X})$, the repeated roots for the second and third series $Pr(r_j^i = r_k^l|\mathbf{X})$, and the roots that appear only in each particular series $Pr(r_j^i \sim C|\mathbf{X})$, where $r_j^i \sim C$ means that r_j^i is restricted to the continuous part of its probability density function. The probabilities that appear in **bold** correspond to the highest marginal posterior probabilities of each particular series. For example, for the first series, there is zero posterior probability that the series was generated from an AR(0) or an AR(1) model. Similarly, for the second series, the posterior probability that its first root (r_1^2) is the same as the first root in series one (r_1^1) is 0.321, while the posterior probability of a "new" root, sampled from the continuous part of the distribution, is 0.590. The probability that the second root of the second series (r_2^2) is new equals 0.931. For the third series, the most likely scenario is the one in which the first root (r_1^3) is the same as the first root in the first series (r_1^1) and the second root (r_2^3) is the same as the second root in the first series (r_1^2). Therefore, from these marginals probabilities, we can conclude that the most likely model is the one in which the roots of the first series are different from zero,

Model	Posterior probability
$(C, C, r_1^1, C, r_1^1, r_2^1)$	0.146
$(C, C, C, C, r_1^1, r_2^1)$	0.122
$(C, C, C, C, r_1^1, r_2^2)$	0.062
$(C, C, r_1^1, C, r_1^1, r_1^2)$	0.060
$(C, C, C, C, r_1^1, r_2^2)$	0.060

Table 2: *Most likely models from exploration of the joint posterior distribution.*

the roots of the second series are different from zero and also different from the roots of the first series, and the roots of the third series are the same as the roots of the first series. This model can be represented in a vector form as $M_1 : (C, C, C, C, r_1^1, r_2^1)$, where the first component in the vector corresponds to the first root of the first series, the second component to the second root in the series and so on. A value of 0 in any of the components indicates a zero root, C indicates a new or continuous root, and r_k^l indicates that the root is a repeated root and that it corresponds to the k-th root of the l-th series.

An easy way of exploring the joint posterior distribution when models are visited more than once, is by clustering all the models in a tree, using some sort of distance between the different models, and then looking at the models that are at zero distances. If there are repeated models then, the most likely models are the models that cluster in the zero distance groups with more members. Table 2 displays the 6 most likely models obtained after exploring the joint posterior distribution. We obtain that the most likely model is given by $M_2 : (C, C, r_1^1, C, r_1^1, r_2^1)$, that is, a model in which the first series has two pairs of roots different from zero, the second series has a common root with the first series and a continuous root and the two roots in the third series are equal to the roots in the first series. Therefore, the model structure that was visited more times by the MCMC sampling scheme, was the correct model. Figure 3 displays the posterior distributions of all the roots conditional on the model $(C, C, r_1^1, C, r_1^1, r_1^2)$. The first two histograms at the top display the posterior distributions for the modulus and wavelength of the first root for series one, two and three. The two histograms at the center show the posterior distributions for the modulus and the wavelength of the second root of series one and two. Finally, the two histograms at the bottom display the posterior distributions of the modulus and wavelength of the second root of the second series.

6 Conclusions

In this paper, we propose a new class of prior distributions for multivariate times series models that follow a vector autoregressive structure with diagonal coefficient matrices. The class naturally addresses issues about model uncertainty and characteristic root

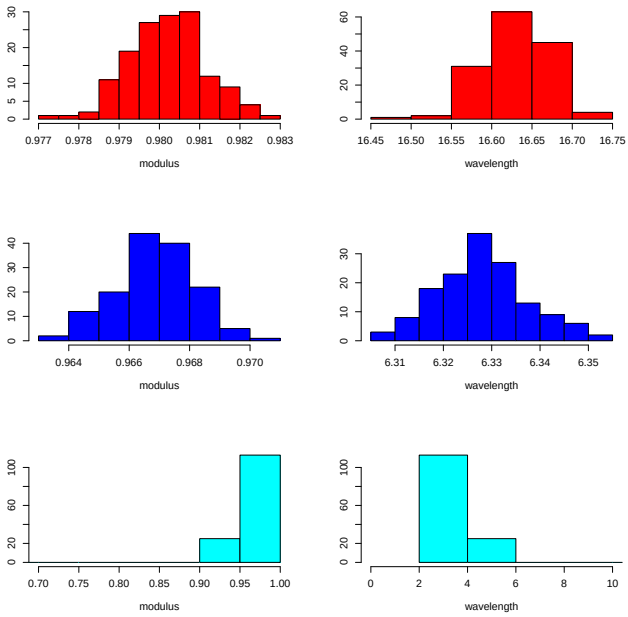

Figure 3: *Posterior distributions for the roots conditional on the model* $(C, C, r_1^1, C, r_1^1, r_1^2)$

structure in a multivariate framework. The structured prior leads to exploration of a very large model space through MCMC simulation. We suggest the use of clustering ideas for more efficient simulation of the posterior distributions of interest.

In this work we assumed that the innovation error covariance matrix Σ is known. This assumption can be relaxed with the use of inverse-Wishart priors. Alternatively, representations of Σ where the matrix elements take simple parametric forms like $\sigma^2\rho^{|i-j|}$, lead to prior specifications of only a few parameters. Reference priors as in Yang and Berger (1994) and the conditionally conjugate prior distributions for covariance matrices presented in Daniels and Pourahmadi (2002) can also be used.

For the case of general VAR processes, i.e., VAR processes with coefficient matrices $\mathbf{\Phi}_j$ values of arbitrary form, it is not trivial to extend the prior structure developed for the diagonal case. The latent processes for each of the scalar components in the multivariate series are defined in terms of the roots of the characteristic polynomial of the VAR process, but in this case, the VAR characteristic polynomial cannot be written as the product of individual characteristic polynomials. This extension deserves more attention and it will be considered for future research.

Bibliography

Aguilar, O., Huerta, G., Prado, R. and West, M. (1999) Bayesian inference on latent structure in time series (with discussion). In *Bayesian Statistics 6* (ed. J.M. Bernardo et al.), pp. 3–26.

Daniels, M. J. and Pourahmadi, M. (2002) Bayesian analysis of covariance matrices and dynamic models for longitudinal data. *Biometrika*, **89,** 553–566.

Gamerman, D. (1997) *Markov Chain Monte Carlo.* London: Chapman & Hall.

Godsill, S.J. and Rayner, P.J.W. (1998) *Digital Audio Restoration: A Statistical Model-Based Approach.* Berlin: Springer-Varlag.

Huerta, G. and West, M. (1999) Priors and component structures in autoregressive time series models. *J. R. Statist. Soc. B.*, **61,** 881–899.

Kitagawa, G. and Gersch, W. (1996) *Smoothness Priors Analysis of Time Series,* Lecture Notes in Statistics, vol. 116. New York: Springer-Verlag.

Krystal, A.D., Prado, R. and West, M. (1999) New methods of time series analysis of non-stationary EEG data: eigenstructure decompositions of time-varying autoregressions. *Clinical Neurophysiology*, **110,** 2197–2206.

Lütkepohl, H. (1993) *Introduction to Multiple Time Series Analysis*, 2nd edn. Heidelberg: Springer-Verlag.

Prado, R. (1998) Latent structure in non-stationary time series. Ph.D. Thesis. Duke University, Durham, NC.

Prado, R. and Huerta, G. (2002) Time-varying autoregressions with model order uncertainty. *Journal of Time Series Analysis*, **23,** 599–618.

West, M. (1997) Time series decomposition. *Biometrika*, **84,** 489–94.

West, M. and Harrison, J. (1997) *Bayesian Forecasting and Dynamic Models*, 2nd edn. New York: Springer-Verlag.

West, M., Prado, R. and Krystal, A.D. (1999) Evaluation and comparison of EEG traces: Latent structure in nonstationary time series. *Journal of the American Statistical Association*, **94,** 1083–1095.

Yang, R. and Berger, J. (1994) Estimation of a covariance matrix using the reference prior. *Annals of Statistics*, **22,** 1195–1211.