Structured priors for multivariate time series

Gabriel Huerta* and Raquel Prado†

Abstract

We present structured prior modeling for 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. 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 analysis are presented.

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

†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.

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

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 that arises in areas such as signal processing. The data usually consist of multiple signals recorded simultaneously, from a system under certain conditions 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 (Huerta and West, 1999b; Aguilar *et al.*, 1999; Godsill and Rayner, 1998; West *et al.*, 1999; Krystal *et al.*, 1999; Kitagawa and Gersch, 1996).

One of the major methodological challenges pursued in this paper is the development of structured priors for vector autoregressions that allow the modeling of uncertainty in the number and form of the latent processes related to each series and permit the expression of prior beliefs on characteristic roots, possibly with unitary and 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 struc-

tured priors; therefore MCMC methods for parameter estimation are necessary. Priors on latent component structure were introduced for univariate AR models in Huerta and West (1999b). In this sense, the models proposed in this paper are an extension to the multivariate framework. Although diagonal VAR could be perceived as of limited use, with the structured priors that we propose here, they form a milestone for the modeling of multivariate time series. For example, in EEG analysis as described in the above references, it is very relevant to determine a probabilistic assessment of common latent structures, something that cannot be obtained with a univariate analysis.

2 Multivariate time series decompositions

In this section, we describe general time series decomposition results for a class of multivariate time series processes. The proposed approach focuses on models that can be written in a multivariate dynamic linear model (MDLM) form. We discuss such results in detail for the case of *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 (1999b), 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 $\mathbf{y}_t = (y_{1,t}, \ldots, y_{m,t})'$, modeled using a 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

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

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 divided 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

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

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

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

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

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

Then, by the univariate AR decomposition result discussed in West (1997), each $z_{i,t,j}$ is a quasi-periodic process 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, $\{\mathbf{x} : (\mathbf{G} - \lambda\mathbf{I}_{mp})\mathbf{x} = \mathbf{0}_{mp}\}$. In order to prove that $\mathbf{G}$ is diagonalizable, we have to check that the geometric multiplicity of each eigenvalue of $\mathbf{G}$ equals its algebraic multiplicity. Let λ be any eigenvalue of $\mathbf{G}$ with algebraic multiplicity $m_{a,\lambda}$. Then, λ is either a real or a complex characteristic reciprocal root of $\mathbf{\Phi}(u)$, i.e. $\lambda = r_j \exp(i\omega_j)$, $\lambda = r_j \exp(-i\omega_j)$ or $\lambda = r_j$ for some j. The geometric multiplicity of λ, $m_{g,\lambda}$ is the dimension of the characteristic subspace of λ, $\{\mathbf{x} \;:\; (\mathbf{G} - \lambda\mathbf{I}_{mp})\mathbf{x} = \mathbf{0}_{mp}\}$. The solutions of the system $(\mathbf{G} - \lambda\mathbf{I}_{mp})\mathbf{x} = 0$, with

$\mathbf{x} = (x_{1,1}, \ldots, x_{1,m}, \ldots, x_{p,1}, \ldots, x_{p,m})$ must satisfy the m equations,

$$\begin{array}{ccccccccc}
(\phi_{1,1} - \lambda)x_{1,1} & + & \phi_{1,2}x_{2,1} & + & \ldots & + & \phi_{1,p}x_{p,1} & = & 0 \\
\phi_{2,1}x_{1,2} & + & (\phi_{2,2} - \lambda)x_{2,2} & + & \ldots & + & \phi_{2,p}x_{p,2} & = & 0 \\
\vdots & + & \ddots & + & \cdots & + & \vdots & \vdots & \vdots \\
\phi_{m,1}x_{1,m} & + & \phi_{m,2}x_{2,m} & + & \ldots & + & (\phi_{m,p} - \lambda)x_{p,m} & = & 0
\end{array}$$

and the set of $mp - m = m(p-1)$ equations,

$$\begin{array}{ccccc}
x_{1,1} & - & \lambda x_{2,1} & = & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
x_{1,m} & - & \lambda x_{2,m} & = & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
x_{p-1,1} & - & \lambda x_{p,1} & = & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
x_{p-1,m} & - & \lambda x_{p,m} & = & 0.
\end{array}$$

Using the last $m(p-1)$ equations we obtain $x_{i,j} = \frac{x_{1,j}}{\lambda^{i-1}}$, for $i = 2, \ldots, p$ and $j = 1, \ldots, m$. Substituting these expressions in the first m equations we obtain that

$$x_{1,j}\left(1 - \phi_{1,j}\left(\frac{1}{\lambda}\right) - \phi_{2,j}\left(\frac{1}{\lambda}\right)^2 - \ldots - \phi_{p,j}\left(\frac{1}{\lambda}\right)^p\right) = 0, j = 1, \ldots, m. \tag{9}$$

Now, $\lambda \neq 0$ has algebraic multiplicity $m_{a,\lambda}$, therefore, λ is a reciprocal root of $m_{a,\lambda}$ characteristic polynomials. Let $j_1, \ldots, j_{m_{a,\lambda}}$ be the series associated to such polynomials. Then, the equations (9) have non trivial solutions x_{1,j_k} for $k = 1, \ldots, m_{a,\lambda}$ and all the other elements of $\mathbf{x}$ can be written as functions of $x_{1,j_k}, k = 1, \ldots, m_{a,\lambda}$. This

implies that $m_{g,\lambda} = m_{a,\lambda}$ for all λ and then $\mathbf{G}$ is diagonalizable, i.e. $\mathbf{G} = \mathbf{BAB}^{-1}$ with $\mathbf{A}$ the diagonal matrix of eigenvalues, or reciprocal characteristic roots, and $\mathbf{B}$ a corresponding matrix of eigenvectors.

3 The prior structure

We extend the priors on autoregressive root structure developed in Huerta and West (1999b), and studied for spectral estimation in Huerta and West (1999a), to the context of vector autoregressions with diagonal matrices of coefficients or DVARs. Some specific aspects of the prior are discussed next. In order to keep the notation as clear as possible, we present the prior distribution for a two-dimensional VAR model. This structure can be generalized for a $\text{VAR}_m(p)$ process.

Assume that we have an m-dimensional series where $m = 2$. 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 $i = 1, \ldots, 2$. 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;
- $U(\cdot|a, b)$ denotes a Uniform distribution over the interval (a, b).

Then, we assume the following prior structure on the component roots of the $m = 2$ series.

-(a) <u>*Priors for real roots.*</u> Let $R_1 = 2$ and $R_2 = 2$ be the maximum number of real roots of the first and second series respectively. Additionally, let r^i_j denote the root j of the series i and $\mathbf{r}^i_{1:R_i}$ all the real roots of series i. A conditional prior structure is proposed, $p(\mathbf{r}^1_{1:R_1}, \mathbf{r}^2_{1:R_2}) = p(\mathbf{r}^1_{1:R_1}) \times p(\mathbf{r}^2_{1:R_2}|\mathbf{r}^1_{1:R_1})$, such that $p(\mathbf{r}^1_{1:R_1}) = \prod_{j=1}^2 p(r^1_j)$ and $p(\mathbf{r}^2_{1:R_2}|\mathbf{r}^1_{1:R_1}) = p(r^2_1|\mathbf{r}^1_{1:R_1}) \times p(r^2_2|\mathbf{r}_{1:R_1}, r^2_1)$. Specifically, we have the following structure for the roots of the first series

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

for $j = 1, 2$ and $g_r(\cdot)$ a continuous density over $(-1, 1)$. Note that by definition the prior is on the stationary region. The mass probability $\pi_{r,0}$ is a prior probability at $r^1_j = 0$. This prior probability at zero allows the modeling of uncertainty in the number

of latent components. Additionally, prior point masses at -1 and 1 can be incorporated to allow the possibility of non-stationary components (see Figure 1). Now, for the roots of the second series we have

$$\begin{aligned} r_1^2|r_1^1, r_2^1 &\sim \pi_{r,0} I_0(r_1^2) + \pi^*_{r,r_1^1} I_{r_1^1}(r_1^2) + \pi^*_{r,r_2^1} I_{r_2^1}(r_1^2) + (1 - \pi_{r,0} - \pi^*_{r,r_1^1} - \pi^*_{r,r_1^2}) g_r(r_1^2) \\ r_2^2|r_1^1, r_2^1, r_1^2 &\sim \pi_{r,0} I_0(r_2^2) + \pi^*_{r,r_1^1} I_{r_1^1}(r_2^2) + \pi^*_{r,r_2^1} I_{r_2^1}(r_2^2) + (1 - \pi_{r,0} - \pi^*_{r,r_1^1} - \pi^*_{r,r_1^2}) g_r(r_2^2), \end{aligned}$$

where π^*_{r,r_j^1} are prior probabilities on the roots of the first series if such roots are different from 0, and have not been sampled already as roots of the second series, i.e., "repeated" roots within the same series are not permitted.

Various choices for $g_r(\cdot)$ can be considered. For instance, the reference prior is the uniform distribution $g_r(\cdot) = U(\cdot| -1, 1)$, i.e., the formal reference prior for the component AR(1) coefficient r_j^i truncated to the stationary region. The $\pi_{r,0}$ and the $\pi^*_{r,\cdot}$ can be considered fixed tuning parameters or alternatively, as it is usually preferred in many applications, they can be treated as hyperparameters to be estimated. In the later case relatively or absolutely uniform priors that can be viewed as non-informative priors should be imposed on these probabilities. Huerta and West (1999b) propose the use of Dirichlet prior distributions for the univariate case.

To illustrate the prior on the real reciprocal roots, we use Figure 1. The first row corresponds to the roots of the first series and on the second row, to the second series. We are assuming that the continuous part of the prior is $U(\cdot| -1, 1)$ and the different probability masses are represented by vertical lines. In the figure, we are also including probability masses at the boundary points, -1 and 1, although we eliminated

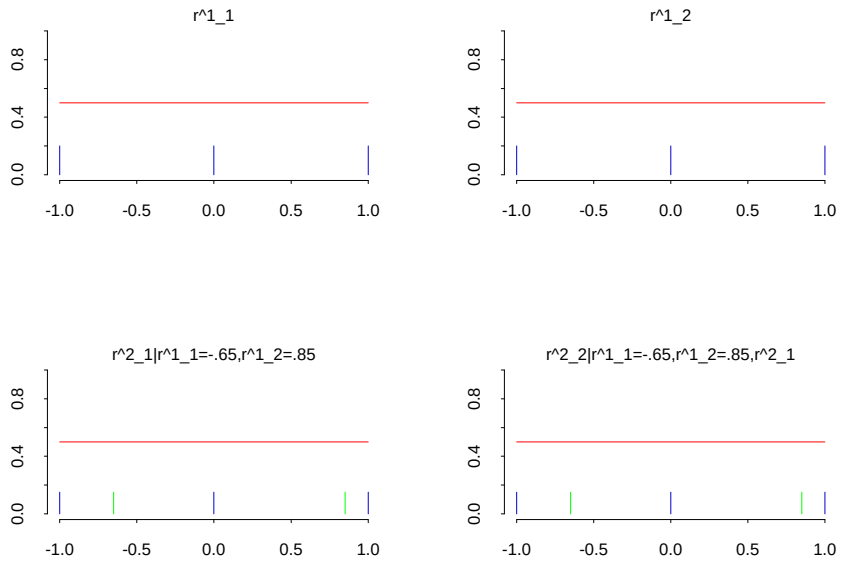

Figure 1: *Prior on real roots.* $m = 2$; $R_1 = R_2 = 2$. $g_r(\cdot) = U(\cdot| -1, 1)$. The vertical lines along the $(-1, 1)$ axis represent the probability masses for each reciprocal root.

this masses for clarity. For the roots of the second series, the prior is conditional on $r_1^1 = -0.65$ and $r_2^1 = .85$ so point masses appear at these two values.

-(b) *Priors for complex roots.* The structure for the complex roots is similar to that proposed for the real roots. Again, it is necessary to specify upper bounds for the maximum number of pairs of complex roots for each series, or equivalently, for the maximum number of quasi-periodic latent processes, and then use a conditional structure. In order to illustrate how this is done, assume for instance that $m = 2$ and $C_1 = C_2 = 2$ are the maximum number of pairs of complex roots of the form $\alpha_j^i = (r_j^i, \lambda_j^i)$, with r_j^i and $\lambda_j^i = 2\pi/\omega_j^i$ the modulus and wavelength of the j-th quasi-periodic process for

series i. Then, a conditional prior structure $p(\boldsymbol{\alpha}^1_{1:C_1}, \boldsymbol{\alpha}^2_{1:C_2}) = p(\boldsymbol{\alpha}^1_{1:C_1}) \times p(\boldsymbol{\alpha}^2_{1:C_2}|\boldsymbol{\alpha}^1_{1:C_1})$, is proposed. The component roots of the first series have an independent prior structure, $p(\boldsymbol{\alpha}^1_{1:C_1}) = p(r^1_1)p(\lambda^1_1)p(r^1_2)p(\lambda^1_2)$ with priors specified over support $0 \leq r^1_j \leq 1$ and $2 < \lambda^1_j < \lambda_u$ for $j = 1, 2$, and a given upper bound λ_u on the wavelengths. Specifically,

$$r^1_j \sim \pi_{c,0} I_0(r^1_j) + (1 - \pi_{c,0}) g_c(r^1_j), \quad \lambda^1_j \sim h(\lambda^1_j),$$

with $h(\lambda^1_j)$ a density over the support $(2, \lambda_u)$ and $g_c(\cdot)$ a continuous density over $(0, 1)$. $\pi_{c,0}$ represents a probability mass at value 0 for the modulus of the root. To address for non-stationary, a point mass at 1 for r^1_j can also be included for the prior (see Figure 2). Similar to the real case, the priors on the AR structure for the complex roots of the second series, α^2_j, are conditional on the root components of the first series and on the complex roots previously sampled for the second series, that is

$$\begin{aligned}
r^2_1|r^1_1, r^1_2 &\sim \pi_{c,0} I_0(r^2_1) + \pi^*_{c,r^1_1} I_{r^1_1}(r^2_1) + \pi^*_{c,r^1_2} I_{r^1_2}(r^2_1) + (1 - \pi_{c,0} - \sum_{j=1}^{2} \pi^*_{c,r^2_j}) g_c(r^2_1) \\
r^2_2|r^1_1, r^1_2, r^2_1 &\sim \pi_{c,0} I_0(r^2_2) + \pi^*_{c,r^1_1} I_{r^1_1}(r^2_2) + \pi^*_{c,r^1_2} I_{r^1_2}(r^2_2) + (1 - \pi_{c,0} - \sum_{j=1}^{2} \pi^*_{c,r^2_j}) g_c(r^2_2) \\
\lambda^2_1|\alpha^1_1, \alpha^1_2 &\sim \sum_{j=1}^{2} I_{r^1_j}(r^2_j) I_{\lambda^1_j}(\lambda^2_1) + [1 - \sum_{j=1}^{2} I_{r^1_j}(r^2_j) I_{\lambda^1_j}(\lambda^2_1)] h(\lambda^2_1) \\
\lambda^2_2|\alpha^1_1, \alpha^1_2, \alpha^2_1 &\sim \sum_{j=1}^{2} I_{r^1_j}(r^2_j) I_{\lambda^1_j}(\lambda^2_2) + [1 - \sum_{j=1}^{2} I_{r^1_j}(r^2_j) I_{\lambda^1_j}(\lambda^2_2)] h(\lambda^2_2).
\end{aligned}$$

Different choices for $g_c(r^i_j)$ and $h(\lambda^i_j)$ can be considered, including uniform priors and margins for λ^i_j based on uniform priors for the corresponding frequency ω^i_j. The default prior is the "component reference prior" (Huerta and West, 1999b) induced by assuming a uniform prior for the implied AR(2) coefficients $2r^i_j \cos(2\pi/\lambda^i_j)$ and $-(r^i_j)^2$ but

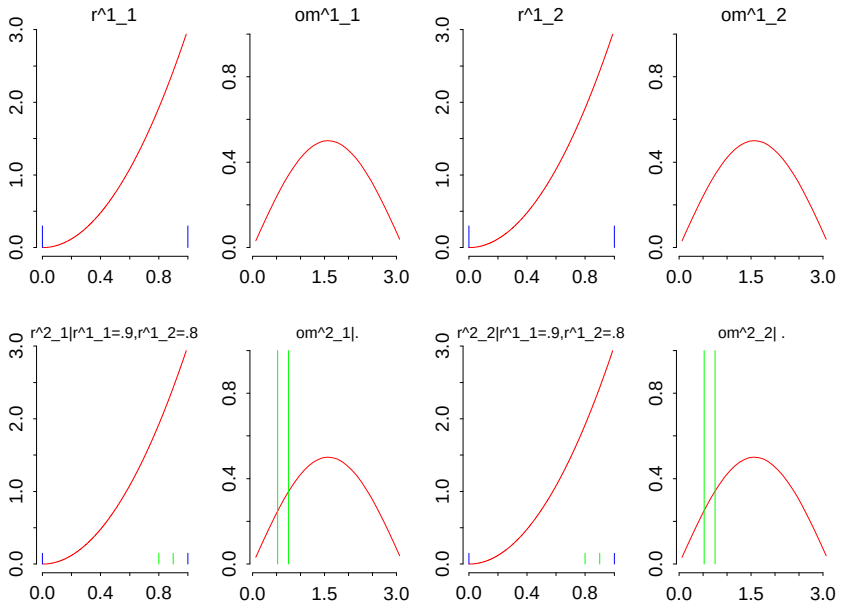

Figure 2: *Prior on Complex roots* $m = 2$; $C_1 = C_2 = 2$; $g_c(\cdot)$ and $h(\cdot)$ are specified with the component reference prior. The figure shows the marginals for r_j^i and ω_j^i. The vertical lines indicate a probability mass.

with finite support for λ_j^i. In addition, as with the real roots, relatively or absolutely continuous priors can be imposed on $\pi_{c,0}$ and $\pi_{c,\cdot}^*$.

In Figure 2 we illustrate the prior for the complex case. The first row presents the marginals for the modulus and frequencies $(r_j^1, \omega_j^1); j = 1, 2$ of the reciprocal roots corresponding to the first series. The second raw shows the marginals for $(r_j^2, \omega_j^2); j = 1, 2$. The continuous part of these marginal densities is defined by the *component reference prior* of Huerta and West (1999b), then $g_c(r_j^i) = Beta(r_j^i|3, 1)$ and $h(\omega_j^i) \propto sin(\omega_j^i)$. For the first row, the vertical lines represent probability masses for the moduli at 0 and at 1. For the second row, the vertical lines also appear at specific values of

the modulus and frequency corresponding to the first series. The prior is conditional on $r_1^1 = 0.9$, $r_1^2 = 0.8$, $\omega_1^1 = 0.52$ and $\omega_2^1 = 0.75$.

Although the DVAR model may be perceived as of limited practical use, it establishes a milestone in prior specifications over the characteristic roots of a VAR model. Additionally, the DVAR models presented here are capable of identifying common latent structure across series, something that has not been addressed in the specialized literature before.

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 3 displays two-dimensional

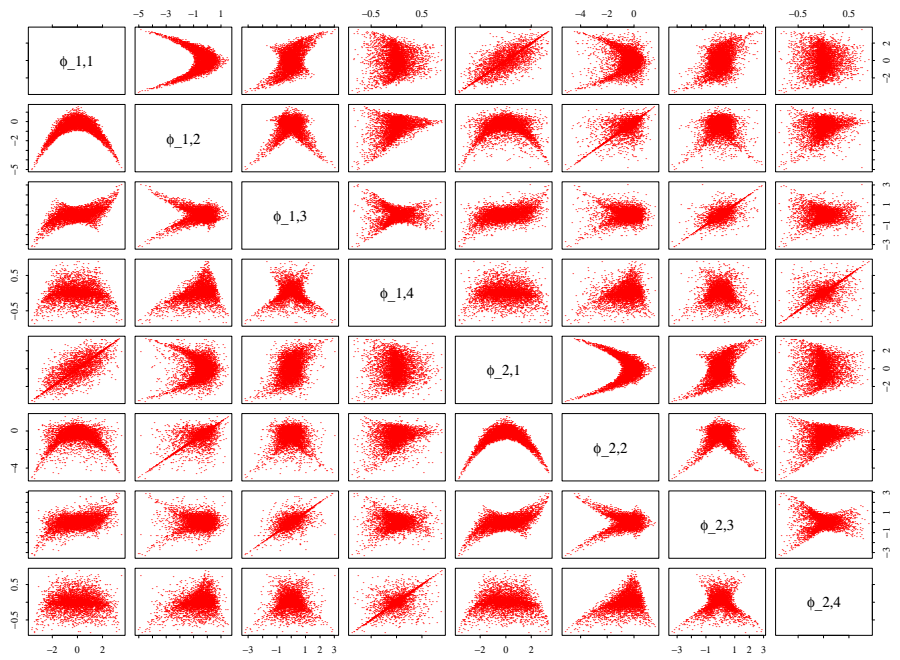

Figure 3: *Samples from the prior for ϕ in a DVAR$_2$(4) model with $R_1 = R_2 = 2$ and $C_1 = C_2 = 1$.*

margins of a sample of 10,000 draws from the prior. The two-dimensional margins of the 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 1999b). 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 3 are contained in this region. Note that 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. The structure of relevant conditional posterior distributions is briefly outlined here.

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 a normal distribution with mean m and variance M truncated to $(-1, 1)$, and the values (m_j^i, M_j^i) and point masses can be easily 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 $z_{t,l} = \prod_{k=1}^{p_l}(1 - \alpha_j^i B) x_{t,l}$ for $l \neq j$. 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, given that the support of $(\phi_{j,1}^i, \phi_{j,2}^i)$ is a bounded region defined by the stationary condition of the process, sampling from the resulting conditional posterior directly is difficult and because of this, following Huerta and West (1999b), 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 (1999b) for the univariate case. However, the number of computations increases considerably when the number of series

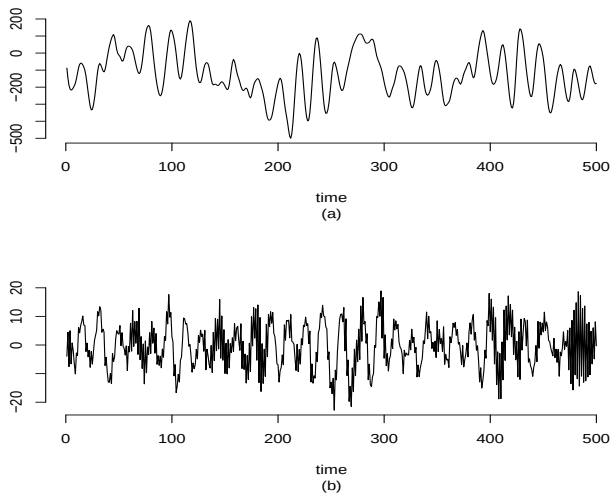

Figure 4: Simulated series. Graphs (a) and (b) correspond to the first and second series.

and/or the model order are large. This will be addressed in the following examples.

5 Examples

Example 1: Figure 4 displays two time series of 500 observations simulated with innovation covariance $\Sigma = 10.0 * \mathbf{I}_3$ and the following latent structure. The first series corresponds to an AR process with one real root with modulus $r_1^1 = 0.98$ and two pairs of complex roots with modulus and wavelengths of $r_2^1 = 0.97$, $r_3^1 = 0.8$ and $\lambda_2^1 = 17.0$, $\lambda_3^1 = 6.0$, respectively. The second series has one common pair of roots with the first series, namely $r_2^2 = 0.97$ and $\lambda_2^2 = 17.0$, another pair characterized by $r_3^2 = 0.8$ and $\lambda_3^2 = 3.0$ and a real root $r_1^2 = -0.98$. Parameter estimation using a prior structure with a maximum of three pairs of complex roots for each series, $C_i = 3$, and one real

real root, $R_i = 1$ for $i = 1, 2$ is done running a reversible jump MCMC algorithm as detailed in the previous section. The results presented here are based on 4,000 samples from the posterior distribution taken after convergence was achieved. The most popular model, expressed in vector form, was $M_1 : (R, 0, C, C, R, 0, \alpha_3^1, C)$, with posterior probability 0.38. In this notation, the first four components of the vector refer to the $R_1 + C_1 = 1 + 3 = 4$ roots of the first series and the last four to the $R_2 + C_2 = 4$ roots of the second series. Then, according to model M_1, the first series has one real root $\alpha_1^1 \sim R$, a zero root $\alpha_2^1 \sim 0$ and two complex roots $\alpha_3^1, \alpha_4^1 \sim C$. The second series has one real root $\alpha_1^2 \sim R$, a zero root $\alpha_2^2 \sim 0$ and two real roots, one of which is a repeated root, i.e., $\alpha_3^2 = \alpha_3^1$, and the other one is a complex root different from the roots of the first series, $\alpha_4^2 \sim C$. Therefore, the posterior mode captures the right structure used to simulate the two series.

Figure 5 shows the histogram of the posterior samples of the real root for the first series (graph (a)) and the histogram of the posterior samples of the real root for the second series (graph (b)), assuming the correct model is M_1. The points in the histograms indicate the posterior means for each case. As seen in the graphs, the model is appropriately estimating the real roots for the two series. Similarly, Figure 6 shows the histograms of the posterior samples of the complex roots for the first and second series. In these graphs we are conditioning on the model structure M_1. Then, panels (a) and (d) display, respectively, the posterior distributions of the modulus and wavelength of the complex root with the highest modulus for the first and second series,

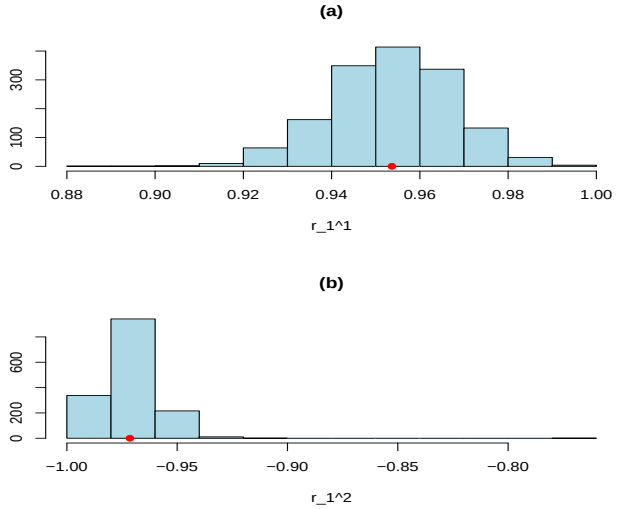

Figure 5: Graphs (a) and (b) display the posteriors of the two real roots for the simulated series.

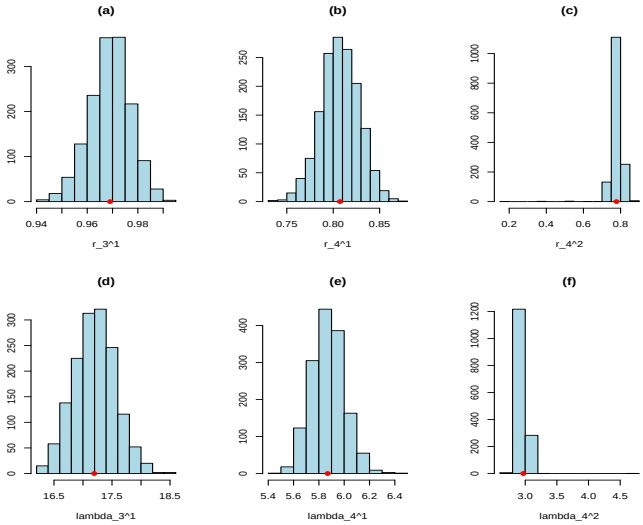

Figure 6: Graphs (a)-(f) display the posteriors for the complex roots of the two simulated series.

i.e., $\alpha_3^1 = \alpha_3^2$. Panels (b) and (e) show the modulus and wavelength of the complex root with the smallest modulus for the first series, α_4^1. Finally, panels (c) and (f) show the modulus and wavelength of the complex root with the smallest modulus for the second series α_4^2. As seen in these graphs, our methodology performs very well in terms of capturing the latent structure present in the simulated data.

The results and figures presented so far are conditional results, i.e., we are looking at the posterior distribution for the parameters conditioning on M_1, the model with the highest posterior probability, being the correct model. It is also possible to report some interesting results obtained by averaging across all possible models. For example, the posterior probability that the two series have a real root different from zero is one: $Pr(\alpha_1^1 \sim R|\mathbf{X}) = Pr(\alpha_1^2 \sim R|\mathbf{X}) = 1.0$. The posterior probability that each series has a zero root is 0.721 for the first series and 0.826 for the second series. Another interesting result is the probability that the second series has a complex root that also appears in the first series. Based on the 4,000 posterior samples we obtain that this probability is 0.6975. In the following example we will discuss some issues related to the exploration of the posterior distribution.

Example 2: A DVAR$_3$(4) with two pairs of complex roots.

Even after MCMC convergence has been achieved, exploring the posterior distribution from a sample is not necessarily trivial when structured priors of the form described in Section 3, or extensions of such priors, are considered for VAR models. We use a simulated example to illustrate some of the difficulties encountered when a

three-dimensional DVAR model was fitted to simulated series that have a small number of latent components. Such difficulties will be greater for models that deal with a large number of series and a large number of latent processes. 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 (7) 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 prior on $\pi^*_{c_i \cdot}$ for each series, while $g_c(\cdot)$ and $h(\cdot)$ are taken as component reference priors as described in Section 3. 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

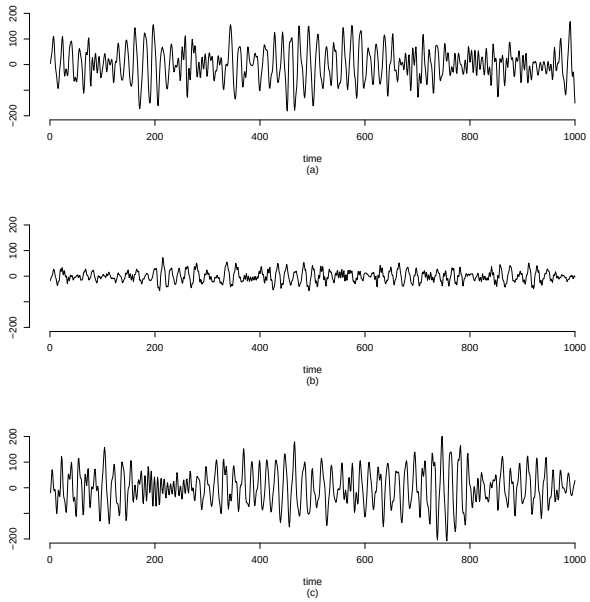

Figure 7: *Simulated series. Graph (a) and (b) correspond to processes with two pairs of complex roots. The series in graph (c) was simulated using the same root structure used to generate the series displayed in (a).*

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, 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

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*

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.*

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.

A possible way of finding models with high joint posterior probabilities in such cases is by means of clustering analysis, following an idea proposed in Bielza *et al.* (1996) and used in Sansó and Müller (1997) in the context of optimal design problems. If a distance between models is defined, then it is possible to produce a cluster tree, cut the tree of model structures at a certain height and consider the sizes of the resulting cluster. 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, in the example above it was possible to cut the tree at height zero and find the cluster with the largest size. This cluster corresponds to the correct model M_2.

6 Conclusions and Extensions

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 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 exploration of the posterior distributions of interest.

In this work we assume 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., VARs with coefficient matrices Φ_j of arbitrary form, it is not trivial to extend the prior structure developed in Section 3.

In particular, extending such hierarchical structured priors in a way that guarantees stationarity of the VAR process is a very difficult task. The latent processes of each of the scalar components in the multivariate series are defined in terms of the roots of the VAR characteristic polynomial, which for general processes cannot be written as the product of individual characteristic polynomials. In future research we will investigate structured priors for triangular VAR processes, i.e., processes for which the matrices of coefficients are either upper or lower triangular matrices, and for transformations of the VAR that lead to a collection of univariate processes that can be fitted separately, such as the transformations proposed by Kitagawa and Gersch (1996).

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.

Bielza, C., Müller, P. and Ríos-Insúa, D. (1996) Monte Carlo methods for decision analysis with applications to influence diagrams. Technical Report 96-07. ISDS, Duke University.

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

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. (1999a) Bayesian inference on periodicities and component spectral structure in time series. *Journal of Time Series Analysis*, **20,** 401–416.

Huerta, G. and West, M. (1999b) 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.

Sansó, B. and Müller, P. (1997) Redesigning a network of rainfall stations. In *Case Studies in Bayesian Statistics, New York*, vol. IV, pp. 383–394. Springer-Verlag.

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.