Structured priors for multivariate time series

Gabriel Huerta* and Raquel Prado†

Abstract

A class of prior distributions for multivariate autoregressive models is presented. This class of priors is built taking into account the latent component structure that characterizes a collection of autoregressive processes. In particular, the state-space representation of a vector autoregressive process leads to the decomposition of each time series in the multivariate process into simple underlying components. These components may have a common structure across the series. A key feature of the proposed priors is that they allow the modeling of such common structure. This approach also takes into account the uncertainty in the number of latent processes, consequently handling model order uncertainty in the multivariate autoregressive framework. Posterior inference is achieved via standard Markov chain Monte Carlo (MCMC) methods. Issues related to inference and exploration of the posterior distribution are discussed. We illustrate the methodology analyzing two data sets: a synthetic data set with quasi-periodic

*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.

latent structure, and seasonally adjusted US monthly housing data consisting of housing starts and housing sold over the period 1965 through 1974.

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

MSC: 62H99, 62P20 and 62P30

Short title: Priors for multiple time series

1 Introduction

We propose a class of prior distributions for vector autoregressive processes with diagonal coefficient matrices or DVAR. The development of these models is motivated by the need of studying multiple series recorded simultaneously in a system under certain conditions. Typically, each one of the multiple time series has an underlying structure, possibly but not necessarily quasi-periodic, that can be adequately described by an autoregressive (AR) model. Some of the components in this underlying structure are usually shared across the multiple series. Data of this kind arise in many applied areas such as signal processing and econometrics. In particular, univariate time series arising in fields that involve seismic recordings, environmental time series, biomedical and speech signals, to mention a few, have such characteristics and have been analyzed in the past using autoregressive models, 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) and Kitagawa and Gersch (1996) present examples in the areas of application mentioned above.

A key component of the prior distributions and methods developed here, is that they model the uncertainty in the number and form of the latent processes related to each univariate series. In addition, these models provide a way to incorporate prior beliefs on the characteristic roots of the AR processes, including unitary and zero roots. Finally, this methodology allows the modeler to consider common latent components across the series, which is a very important feature in many applications. For instance,

the univariate analyses of several electroencephalogram series recorded on a patient under ECT (a treatment of major depression) presented in West *et al.* (1999) and Krystal *et al.* (1999), suggest that the multiple series are characterized by a common underlying structure with two, or possible three, quasi-periodic components. In this particular application it is relevant to obtain a probabilistic assessment of such common latent structure. This can only be done through a multivariate analysis of the traces in the which the possibly common underlying structure across series is explicitly included in the prior distribution.

Although DVAR models could be perceived as models with very limited practical use, when combined with the structured priors presented here, they form a flexible class of models that can be used to search for latent structure in multiple time series from a multivariate perspective. Therefore, the implementation of DVAR models constitutes a first major step towards developing more sophisticated multivariate models that can be useful in analyzing very complex multivariate data, such as the EEG series considered in West *et al.* (1999) and Krystal *et al.* (1999).

Priors on latent component structure were introduced for univariate AR models in Huerta and West (1999b). In this sense, the models proposed here are a multivariate extension to the models developed by Huerta and West. As in the univariate case, posterior inference is achieved via customized MCMC methods. However, additional computational difficulties arise in the multivariate framework when considering many multiple series with a rich latent component structure. In particular, the exploration

of the posterior distribution may be a difficult task. This is highlighted in one of the examples presented in Section 5. Some alternatives for exploring and summarizing the posterior distribution are investigated.

The paper is organized as follows. Section 2 summarizes the multivariate decomposition results that motivate the development of the structured priors for DVAR models. Section 3 defines the prior structure in detail and discusses some aspects of assuming such structure with some examples. Section 4 describes the MCMC methodology to achieve posterior inference. Section 5 illustrates the methodology with two examples and finally, Section 6 presents a discussion and points towards future extensions.

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 autoregressive (DVAR) models. We begin by revisiting the developments on multivariate time series decompositions presented in Prado (1998), including some extensions that handle more general models.

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 $\mathbf{x}_t$ as follows

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

with $\mathbf{F}_i$ the i-th column of the matrix $\mathbf{F}$. We now show that each component $x_{i,t}$ can be written as a sum of 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 $\mathbf{G}_t = \mathbf{B}_t \mathbf{A}_t \mathbf{B}_t^{-1}$. 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. This is, 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$_m(p)$

Consider the case of an m-dimensional time series process $\mathbf{x}_t$ that follows a VAR$_m(p)$

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

where $\boldsymbol{\Phi}_j$ for $j = 1, \ldots, p$ 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.

Any $\text{VAR}_m(p)$ process can be written in the MDLM form (1), with $d = mp$, $\boldsymbol{\nu}_t = 0$, 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

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

Here 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

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

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

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

i.e. they are the reciprocal roots of the characteristic polynomial given by $\boldsymbol{\Phi}(u) = det(I_m - \boldsymbol{\Phi}_1 u - \ldots - \boldsymbol{\Phi}_p u^p)$. Therefore, $\mathbf{x}_t$ is stable, and consequently stationary, if the eigenvalues of $\mathbf{G}$ have modulus less than one (see for instance Lütkepohl, 1993). 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 (2), (3) and the fact that $\mathbf{K}_{i,t} = \mathbf{I}$ for all i, j we have

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

Now, following the univariate AR decomposition results discussed in West (1997), we obtain that 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$_m$(p).* Suppose that we have a VAR$_m(p)$ 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

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

with $\mathbf{\Phi}^i(u)$ being the characteristic polynomial of series i. In other words, $\mathbf{\Phi}(u)$ 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, this 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. We now prove 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, $\{\mathbf{x} : (\mathbf{G} - \lambda \mathbf{I}_{mp})\mathbf{x} = \mathbf{0}_{mp}\}$. 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 additional 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 have that

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

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 DVAR models. We also discuss some specific aspects of the prior. 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 easily generalized for a $\text{VAR}_m(p)$ process.

Assume that we have an m-dimensional series with $m = 2$. 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;

- $\pi_{r,-1}, \pi_{r,0}$ and $\pi_{r,1}$ denote the prior probabilities that a given real root takes the values $-1, 0$ and 1 respectively. Similarly, $\pi_{c,0}$ and $\pi_{c,1}$ denote the prior probabilities that a given complex root takes the values 0 and 1 respectively.

- π^*_{r,r_j^i} denotes the prior probability that a real root of a series different from i, takes the value r_j^i conditional on r_j^i being different from $0, -1$ and 1, and also different from any of the roots that have already been sampled for such series.

This is, "repeated" roots within the same series are not permitted. Similarly, π^*_{c,r^i_j} denotes the prior probability that the modulus of a complex root of a series different from i takes the value r^i_j, conditional on r^i_j being different from 0 and 1 and also different from any of the roots that have already been sampled for such series.

- $\mathbf{r}^i_{1:j} = \{r^i_1, \ldots, r^i_j\}$; $\boldsymbol{\lambda}^i_{1:j} = \{\lambda^i_1, \ldots, \lambda^i_j\}$; $\boldsymbol{\alpha}^i_{1:j} = (\mathbf{r}, \boldsymbol{\lambda})^i_{1:j} = \{(r^i_1, \lambda^i_1), \ldots, (r^i_j, \lambda^i_j)\}$;
- $I_y(z)$ 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) and $Beta(\cdot|a, b)$ denotes a Beta distribution with parameters a and b.

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

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

for $j = 1, 2$ and $g_r(\cdot)$ a continuous density over $(-1, 1)$. 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, $\pi_{r,-1}$ and $\pi_{r,1}$, are incorporated to allow non-stationary components (see Figure 1). Now, for the roots of the second series we have

$$\begin{aligned} r_1^2|r_1^1, r_2^1 &\sim \pi_{r,-1}I_{-1}(r_1^2) + \pi_{r,0}I_0(r_1^2) + \pi_{r,1}I_1(r_1^2) + \pi^*_{r,r_1^1}I_{r_1^1}(r_1^2) + \pi^*_{r,r_2^1}I_{r_2^1}(r_1^2) + \\ &\quad (1 - \pi_{r,-1} - \pi_{r,0} - \pi_{r,1} - \pi^*_{r,r_1^1} - \pi^*_{r,r_2^1})g_r(r_1^2) \\ r_2^2|r_1^1, r_2^1, r_1^2 &\sim \pi_{r,-1}I_{-1}(r_2^2) + \pi_{r,0}I_0(r_2^2) + \pi_{r,1}I_1(r_2^2) + \pi^*_{r,r_1^1}I_{r_1^1}(r_2^2) + \pi^*_{r,r_2^1}I_{r_2^1}(r_2^2) + \\ &\quad (1 - \pi_{r,-1} - \pi_{r,0} - \pi_{r,1} - \pi^*_{r,r_1^1} - \pi^*_{r,r_2^1})g_r(r_2^2), \end{aligned}$$

where π^*_{r,r_j^1} are prior probabilities on the roots of the first series if such roots are different from $0, \pm 1$, and have not been sampled already as roots of the second series.

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 prior masses $\pi_{r,\cdot}$ and $\pi^*_{r,\cdot}$ 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 those of 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

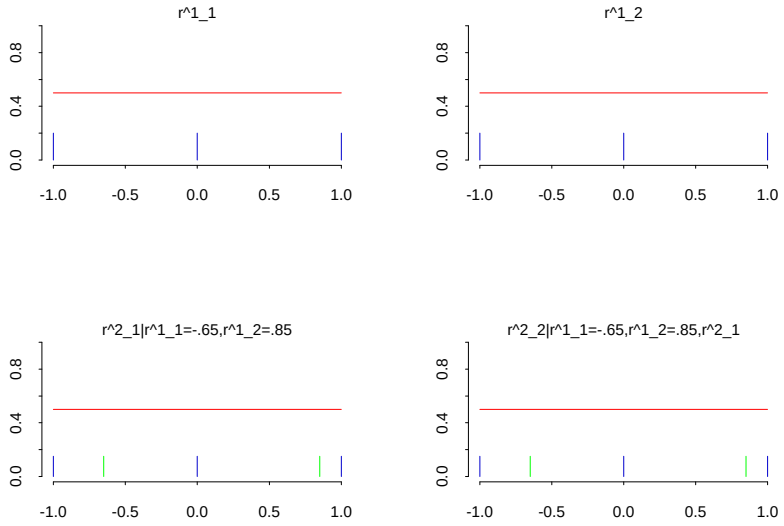

Figure 1: *Prior on real roots.* $m = 2$; $R_1 = R_2 = 2$. $g_r(\cdot) = U(\cdot|-1, 1)$. The vertical lines along the $(-1, 1)$ axis represent the probability masses for each reciprocal root.

including probability masses at the boundary points, -1 and 1. For the roots of the second series, the prior is conditional on $r_1^1 = -0.65$ and $r_2^1 = .85$ and 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,

$$r^1_j \sim \pi_{c,0} I_0(r^1_j) + \pi_{c,1} I_1(r^1_j) + (1 - \pi_{c,0} - \pi_{c,1}) g_c(r^1_j), \quad \lambda^1_j \sim h(\lambda^1_j),$$

with $h(\lambda^1_j)$ a density over the support $(2, \lambda_u)$ and $g_c(\cdot)$ a continuous density over $(0, 1)$. Again, $\pi_{c,0}$ and $\pi_{c,1}$ represent probability masses at values 0 and 1 for the modulus of the root. 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

$$\begin{aligned}
r^2_1|r^1_1, r^1_2 &\sim \pi_{c,0} I_0(r^2_1) + \pi_{c,1} I_1(r^2_1) + \pi^*_{c,r^1_1} I_{r^1_1}(r^2_1) + \pi^*_{c,r^1_2} I_{r^1_2}(r^2_1) + \\
&\quad (1 - \pi_{c,0} - \pi_{c,1} - \sum_{j=1}^{2} \pi^*_{c,r^1_j}) g_c(r^2_1) \\
r^2_2|r^1_1, r^1_2, r^2_1 &\sim \pi_{c,0} I_0(r^2_2) + \pi_{c,1} I_1(r^2_2) + \pi^*_{c,r^1_1} I_{r^1_1}(r^2_2) + \pi^*_{c,r^1_2} I_{r^1_2}(r^2_2) + \\
&\quad (1 - \pi_{c,0} - \pi_{c,1} - \sum_{j=1}^{2} \pi^*_{c,r^1_j}) g_c(r^2_2) \\
\lambda^2_1|\alpha^1_1, \alpha^1_2 &\sim \sum_{j=1}^{2} I_{r^1_j}(r^2_j) I_{\lambda^1_j}(\lambda^2_1) + [1 - \sum_{j=1}^{2} I_{r^1_j}(r^2_j) I_{\lambda^1_j}(\lambda^2_1)] h(\lambda^2_1) \\
\lambda^2_2|\alpha^1_1, \alpha^1_2, \alpha^2_1 &\sim \sum_{j=1}^{2} I_{r^1_j}(r^2_j) I_{\lambda^1_j}(\lambda^2_2) + [1 - \sum_{j=1}^{2} I_{r^1_j}(r^2_j) I_{\lambda^1_j}(\lambda^2_2)] h(\lambda^2_2).
\end{aligned}$$

Different choices for $g_c(r^i_j)$ and $h(\lambda^i_j)$ can be considered, including uniform priors and marginals 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

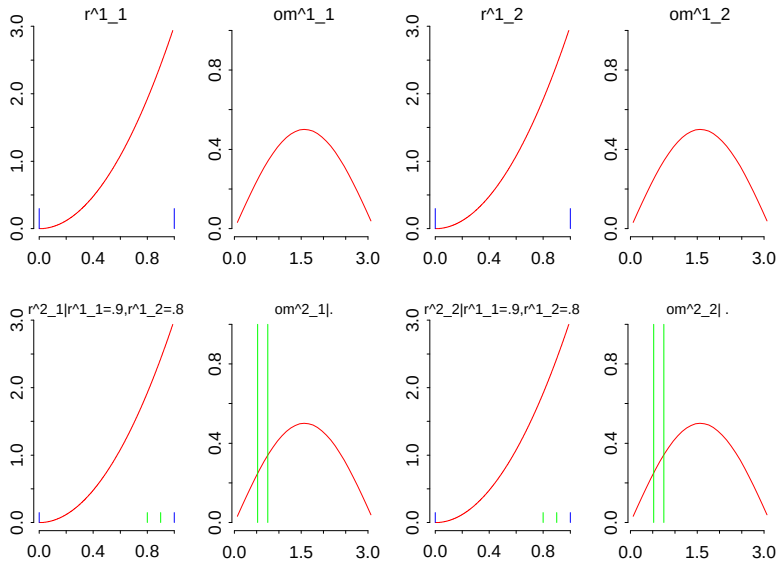

Figure 2: *Prior on complex roots* $m = 2$; $C_1 = C_2 = 2$; $g_c(\cdot)$ and $h(\cdot)$ are specified with the component reference prior. The figure shows the marginals for r_j^i and ω_j^i. The vertical lines indicate a probability mass.

by assuming a uniform prior for the implied AR(2) coefficients $2r_j^i \cos(2\pi/\lambda_j^i)$ and $-(r_j^i)^2$ but with finite support for λ_j^i. In addition, as with the real roots, relatively or absolutely continuous priors can be imposed on $\pi_{c,\cdot}$ 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 of the reciprocal roots of the first series (r_j^1, ω_j^1) for $j = 1, 2$. The second row shows the marginals for the reciprocal roots of the second series, (r_j^2, ω_j^2) for $j = 1, 2$. 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 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 $\omega_2^1 = 0.75$.

This prior structure discussed above can be generalized for the case of m series and arbitrary upper bounds R_i and C_i for $i = 1, \ldots, m$ (Huerta and Prado, 2002). The rich and flexible prior structure placed on functions of the parameters that define the DVAR lead to a models that are capable of identifying common latent structure across multiple series from a multivariate perspective.

3.1 Some aspects of implied prior structure

The prior specified on the reciprocal characteristic roots structure induces priors, of complicated mathematical forms, on the standard linear autoregressive parameters $\phi_{i,k}$, for $i = 1, \ldots, m$ and $k = 1, \ldots, p$. For instance, consider a $VAR_2(5)$ model that allows exactly one real component in each series $R_1 = R_2 = 1$, and two quasi-periodic components in each series, $C_1 = C_2 = 2$. Some of the point masses are set to zero by making $\pi_{r^1,0} = \pi_{r^2,0} = \pi_{c^1,0} = \pi_{c^2,0} = 0$, $\pi_{r^1,-1} = \pi_{r^2,-1} = 0$ 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 probabilities of equal roots, i.e. on the weights $\pi_{r^i,\cdot}$ and $\pi_{c^i,\cdot}$, which are not equal to zero. We explore the implied prior structure on the ten AR coefficients $\boldsymbol{\phi} = (\phi_{1,1}, \ldots, \phi_{1,5}, \phi_{2,1}, \ldots, \phi_{2,5})$ via simulation: for a random draw from the prior in the root parameterization, we can compute the corresponding value of $\boldsymbol{\phi}$ by polynomial multiplication. Figure 3 displays

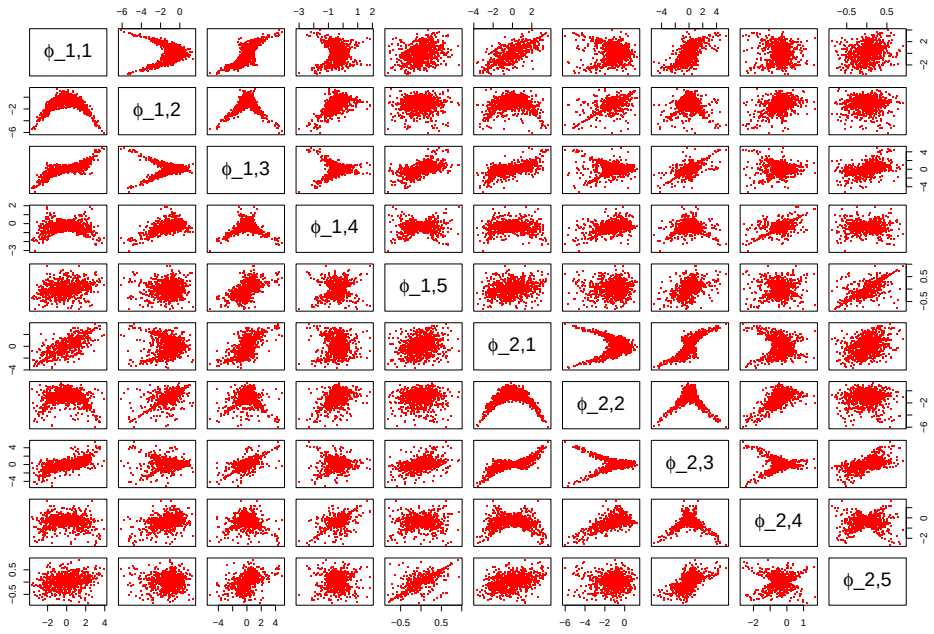

Figure 3: *Samples by pairs from the prior of* ϕ *in a* $DVAR_2(5)$ *model with* $R_1 = R_2 = 1$ *and* $C_1 = C_2 = 2$.

two-dimensional marginals for a sample of 1000 draws from the prior.

The two-dimensional marginals of the AR coefficients $\boldsymbol{\phi}_1 = (\phi_{1,1}, \ldots, \phi_{1,5})$ and $\boldsymbol{\phi}_2 = (\phi_{2,1}, \ldots, \phi_{2,5})$ appear in the five by five diagonal picture blocks. The two five by five 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. The induced priors over $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ are not uniform.

In Figure 4 we show the univariate marginals for each of the model coefficients. The first row shows the marginals for $\boldsymbol{\phi}_1 = (\phi_{1,1}, \ldots, \phi_{1,5})$ and the second row shows the marginals for $\boldsymbol{\phi}_2 = (\phi_{2,1}, \ldots, \phi_{2,5})$. This is a *shrinkage* prior in the sense that as the

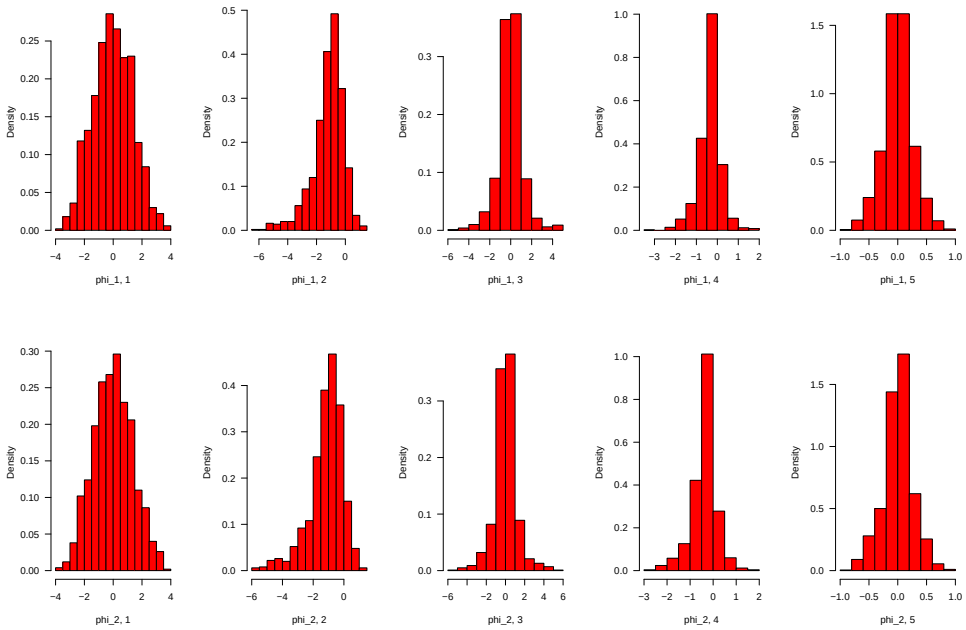

Figure 4: *Univariate margins, prior for ϕ in a DVAR$_2$(5) model with $R_1 = R_2 = 1$ and $C_1 = C_2 = 2$.*

lag of the AR coefficient increases, the prior mass is more concentrated around zero. The marginal prior distribution for each AR coefficient is, in general, not symmetric.

4 Posterior structure in DVAR models

Posterior and predictive calculations are obtained via Markov chain Monte Carlo (MCMC) simulation methods. We briefly outline the structure of relevant conditional posterior distributions.

Assume we have m series. Let $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, with $\mathbf{x}_t = (x_{1,t}, \ldots, x_{m,t})'$, be the observed m-dimensional time series vector. Given the maximum model order p,

with $p = \max\{p_1, \ldots, p_m\}$, 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 that Σ and $\mathbf{X}_0$ are 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 elements elements of $\boldsymbol{\alpha}$ with $\boldsymbol{\xi}$ removed. The MCMC method used to obtain samples from the posterior distribution follows a standard Gibbs sampling format, specifically

- for each $i = 1, \ldots, m$,

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

    2. sample the complex roots individually from $p(\alpha_j^i|\boldsymbol{\alpha}\backslash\alpha_j^i, \mathbf{X}, \mathbf{X}_0, \Sigma)$, for each $j = 1, \ldots, C_i$.

Specific details about how to sample from the distributions in steps 1. and 2. of the MCMC algorithm are given below.

1. *Conditional distributions for real roots.* Assume that we want to obtain a sample from the conditional distribution $p(r_j^i|\boldsymbol{\alpha}\backslash r_j^i, \mathbf{X}, \mathbf{X}_0, \Sigma)$, for some series i and some j. Given all the other model parameters and the DVAR structure, the likelihood function for r_j^i provides a normal kernel. Therefore, under a mixture prior of the form previously described in Section 3, this leads to the mixture posterior

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

Here $N_t(\cdot|m, M)$ denotes the density of a normal distribution with mean m and variance M truncated to $(-1, 1)$. The values (m_j^i, M_j^i) and the point masses can be easily computed. This mixture posterior is easily sampled with direct simulation of the truncated normal by c.d.f. inversion.

2. *Conditional for complex roots.* For each series i, the index j, with $j = 1, \ldots, C_i$, identifies a 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, $\boldsymbol{\alpha}\backslash(r_j^i, \lambda_j^i)$. Given $\boldsymbol{\alpha}\backslash(r_j^i, \lambda_j^i)$ and $\mathbf{X}$ we can directly compute the filtered time series as $z_{t,l} = \prod_{k \in A_j^i}(1 - \alpha_k^l B)x_{t,l}$ 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. Because of this and following Huerta and West (1999b), we use a reversible jump Markov chain Monte Carlo step.

The structure of the MCMC algorithm in the multivariate 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 and/or the model orders are large. This issue will be addressed in the following examples.

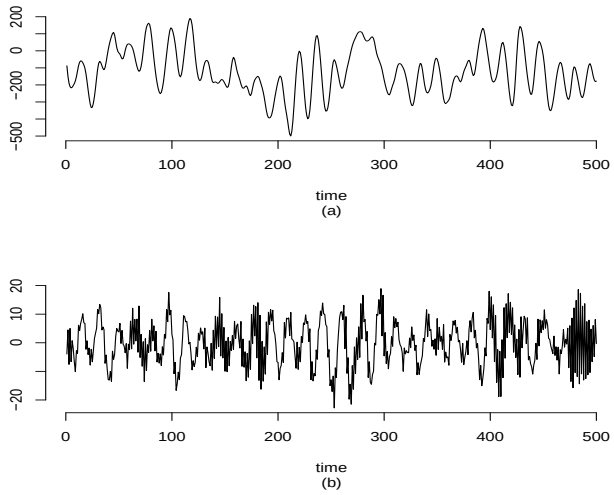

Figure 5: *Simulated series.* Graphs (a) and (b) correspond to the first and second series.

5 Examples

5.1 Analysis of synthetic data

Figure 5 displays two time series of 500 observations simulated with innovation covariance matrix $\Sigma = 10.0 * \mathbf{I}_3$ and the following latent structure. The first series was generated from 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 $C_i = 3$ and one real real root $R_i = 1$ per

series is achieved via the reversible jump MCMC algorithm detailed in the previous section. The prior masses for the roots on the stationary boundary were set to zero. Discrete uniform priors were used for the prior masses of the roots lying in the stationary region. In addition, $g_r(\cdot), g_c(\cdot)$ and $h(\cdot)$ were taken as component reference priors. The results presented here are based on 4,000 samples from the posterior distribution taken after convergence was achieved, following a burn-in period of 10,000 iterations.

Exploring the posterior distribution involves thinking about the possible models that may result from considering the priors described in Section 3. We use a vectorial notation to denote the structure of a given model. For instance, in this example, a possible model structure is $(R_1^1, 0, C_2^1, C_3^1; 0, C_2^1, C_2^2, C_3^2)$. In this notation the first four components of the vector refer to the roots of the first series, while the last four refer to the roots of the second series. This is, for the first series, the first root is a real root different from zero, the second root is a zero root and the third and fourth roots are complex and different from zero. Similarly, for the second series, the first root is a zero root, the second one is a complex root equal to the third root of the first series and the third and fourth components are complex roots different from zero and also different from any of the complex roots of first series. Note that in this example the number of possible models is large, considering that we have a small number of series ($m = 2$) and relatively small model orders (the maximum model order per series is 7). For a single time series with a maximum of one real root and three complex roots, we have a total number of 8 possible models. When a second series with similar structure is added,

the number of possible models increases enormously due to the fact that the roots of the first series can also appear in the second series if common latent structure is shared by the two series. For instance, considering only the models in which all the roots are distinct, or in which only the real root can be repeated and all the complex roots are distinct, we get 80 models. It is easy to see that the total number of models that can be considered in cases where several series with a rich latent component structure have to be analyzed is very large.

One possible way to explore the posterior distribution is by looking at the results marginally. In this example we obtain that the terms with the highest marginal posteriors and their corresponding probabilities are

$$Pr(R_1^1 \sim R|\mathbf{X}) = 1.000, \quad Pr(C_1^1 = 0|\mathbf{X}) = 0.721,$$

$$Pr(C_2^1 \sim C|\mathbf{X}) = 1.000, \quad Pr(C_3^1 \sim C|\mathbf{X}) = 1.000,$$

for the first series and

$$Pr(R_1^2 \sim R|\mathbf{X}) = 1.000, \quad Pr(C_1^2 = 0|\mathbf{X}) = 0.826,$$

$$Pr(C_2^2 = C_2^1|\mathbf{X}) = 0.526, \quad Pr(C_3^2 \sim C|\mathbf{X}) = 0.710,$$

for the second series. Therefore, if we have to select a model structure based on these marginal posterior results we would choose model $M : (R_1^1, 0, C_2^1, C_3^1; R_1^2, 0, C_2^1, C_3^2)$. Then, according to model M, the first series is characterized by one real root, a zero root and two complex roots, while the second series has a real root, a zero root and two complex roots, one of which is a repeated root from the first series.

In this example, the model built from the terms with the highest marginal posteriors is the correct model, since it captures the structured used to simulate the two series. However, this is not necessarily the case in all the applications (see Huerta and Prado, 2002) and so, it is important to look for the model with the highest joint posterior probability. A good way of finding models with high joint posterior probabilities 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 and the models of the resulting cluster. In this case it was possible to follow this idea, cut the tree at zero height, since various models were visited several times, and find the cluster with the largest size. The following three models were the most likely models obtained after exploring the joint posterior distribution

$$\begin{aligned} Pr(R_1^1, 0, C_2^1, C_3^1; R_1^2, 0, C_2^1, C_3^2) &= 0.377 \\ Pr(R_1^1, C_1^1, C_2^1, C_3^1; R_1^2, 0, C_2^1, C_3^2) &= 0.192 \\ Pr(R_1^1, 0, C_2^1, C_3^1; R_1^2, 0, 0, C_3^2) &= 0.092. \end{aligned}$$

Again, the most popular model was the correct model $M : (R_1^1, 0, C_2^1, C_3^1; R_1^2, 0, C_2^1, C_3^2)$ with posterior probability 0.38.

Figure 6 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. The points in the

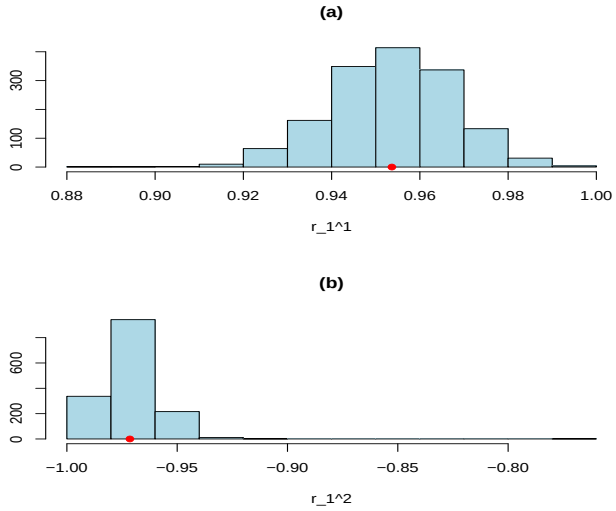

Figure 6: Graphs (a) and (b) display the posteriors of the two real roots for the simulated series.

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

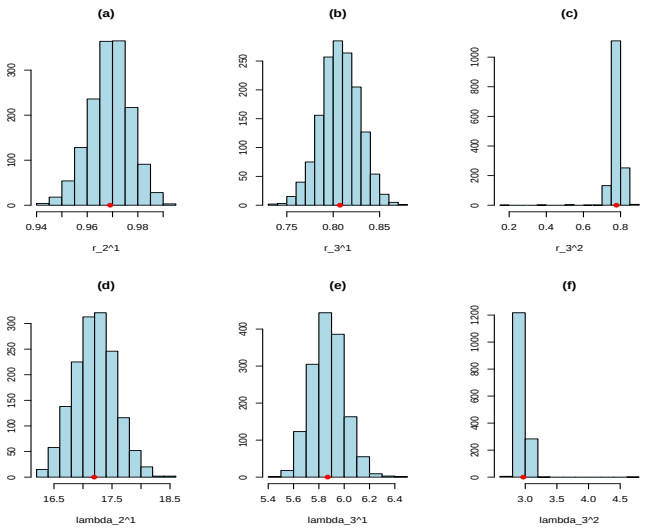

Figure 7: Graphs (a)-(f) display the posteriors for the complex roots of the two simulated series.

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, 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 real roots different from zero is one: $Pr(R_1^1 \neq 0 \ \ \& \ \ R_2^1 \neq 0|\mathbf{X}) = 1.0$. 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.

5.2 Analysis of US monthly housing data

We now consider an application of our methodology for a situation with $m = 2$ time series. This is for the seasonally adjusted U.S. Monthly Housing Data observed during the period January 1965 through December 1974 and where each series consists of 120 observations. The first series is formed by seasonally adjusted housing start values while the second series is formed of housing sold values. The deseasonalized data are obtained by subtracting the monthly average seasonal values from the original series. The series are identified with an AR(2) model (Reinsel, 1997) and so, we selected $C_1 = C_2 = 0$ and $R_1 = R_2 = 2$. Based on 10,000 posterior samples, we present the histograms for the ordered real and reciprocal roots of each series in Figure 8.

Our MCMC produces posterior samples of the roots that are essentially exchangeable. Graphical summaries of these posterior samples have the same characteristics and are not distinguishable. This is a consequence of our priors being invariant to permutations of the roots. A priori and for any given series, the roots are treated as independent variables and so, they are invariant to permutations of the indexes assigned to each root. For posterior analysis, an ordering of the roots needs to be imposed for identifiability and since all the roots considered in this example are real, a natural ordering is $r_1^j < r_2^j$ for $j = 1, 2$, with $j = 1$ assigned to the housing start values and $j = 2$ for the housing sold values. The histograms show that both series are driven by a unit root process with high probability. In fact, the posterior probability that $r_2^1 = 1$ is 0.9295 and the posterior probability that $r_2^2 = 1$ is 0.8023. Additionally, the roots

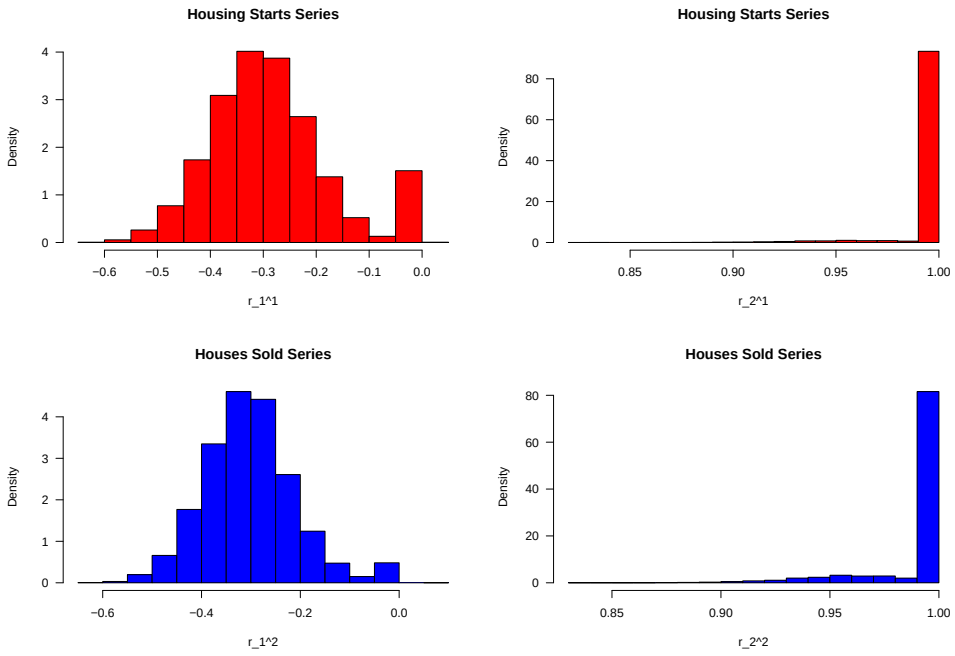

Figure 8: *U.S. Monthly Housing Data:* Posterior histograms of ordered roots for DVAR model with $C_1 = C_2 = 0$, $R_1 = R_2 = 2$.

Figure 9: *U.S. Monthly Housing Data:* (left) Housing Starts Series and latent components associated to r_2^1 and r_1^1. (Right) Housing Sold Series and latent components associated to r_2^2 and r_1^2.

labeled r_1^j have a very similar posterior density function. The posterior probability that $r_1^1 = r_1^2$ is 0.7671. By switching the labels of the series ($j = 1$ housing sold and $j = 2$ housing start), we obtained the same posterior structure for the roots, showing that our models are also invariant to the ordering of the vector time series.

The posterior means of the AR(2) coefficients for each series are given by $\boldsymbol{\phi}_1 = (1.29, -29)$ and $\boldsymbol{\phi}_2 = (1.30, -.3)$. Using these posterior means we obtained the decomposition of latent structure for each of the observed processes. The resulting decompositions and the observed time series are presented in Figure 9.

The components labeled $R1$ are associated to the largest roots r_2^1 and r_2^2. Clearly

these components have the same amplitude of the data and the main characteristics of a random walk process so these components represent a "common trend series". The components labeled $R2$ correspond to the roots r_1^1 and r_1^2 that are not unit and common to both series. These components exhibit the pattern of a stationary AR(1) with a low modulus. Furthermore, Figure 6.1 in Reinsel (1997), Chapter 6, shows a cointegrated process for these data using a $\text{VAR}_2(1)$ formulated via canonical correlations. Our components labeled $R2$ resemble the transformed series showed by Reinsel and correspond to the stationary latent component underlying the two series.

6 Conclusions and Extensions

A new class of prior distributions for multivariate times series models that follow a vector autoregressive structure with diagonal coefficient matrices is presented here. This class of priors naturally incorporates model order uncertainty and characteristic root structure in a multivariate framework.

Vector autoregressions with upper triangular or lower triangular matrices of coefficients, as the DVAR models, have characteristic polynomials that can be written as the product of polynomials associated with each individual series. In future research we will investigate the use of structured priors for triangular VAR models. It is natural to use the priors developed here for the coefficients that lie on the diagonal of the triangular VAR coefficient matrices, since these parameters define the latent structure of the individual series. In addition, a variety of priors can be considered for the coeffi-

cients that lie off the diagonal in triangular VAR processes. Such coefficients model the dependence of one particular series and the lagged values of the rest of the series. So for example, a prior structure with spikes at zero could be used to allow the inclusion or exclusion of such coefficients. If a coefficient is included, a continuous prior can be specified.

For general VAR processes 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 prior structure 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 characteristic polynomial, which for general VAR processes cannot be written as the product of individual characteristic polynomials. In connection with this, transformations of the VAR leading to a collection of univariate processes that can be fitted separately, such as the transformations proposed by Kitagawa and Gersch (1996), will be considered in future extensions of this work.

One of the assumptions made here was that of the innovation error covariance matrix Σ being 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 such as $\sigma^2\rho^{|i-j|}$, 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 explored.

Finally, the proposed structured prior leads to exploration of a very large model space through MCMC simulation. The use of clustering ideas for more efficient exploration of the posterior distributions of interest is initially investigated here. We expect to further investigate this issue by defining distances between models. Such distances can be defined in terms of the closeness of the latent structures of the different models being considered. Models with different structures which may be roughly equivalent in terms of the dominant latent components, or equivalent in terms of predictive performance, would belong to the same class of models. The classes can be defined in terms of a given distance. Then, it would be possible to, for example, choose the most parsimonious model within a particular class in order to describe the structure of the series or for predictive purposes.

Acknowledgments

We wish to express our gratitude to the Associate Editor and to the referee for their comments on a previous version of this paper.

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 Prado, R. (2002) Exploring common structure in multiple time series via structured priors for autoregressive processes. Technical Report. , AMS, UC Santa Cruz and Department of Mathematics and Statistics, University of New Mexico.

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.

Reinsel, G.C. (1997) *Elements of Multivariate Time Series Analysis*, 2nd edn. New York: Springer-Varlag.

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.