Priors and Component Structures in Autoregressive Time Series Models

Gabriel Huerta and Mike West
ISDS, Duke University
Durham, NC 27708-0251. USA
http://www.isds.duke.edu

Abstract

New approaches to prior specification and structuring in autoregressive time series models are introduced and developed. We focus on defining classes of prior distributions for parameters and latent variables related to latent components of an autoregressive model for an observed time series. These new priors naturally permit the incorporation of both qualitative and quantitative prior information about the number and relative importance of physically meaningful components that represent low frequency trends, quasi-periodic sub-processes, and high frequency residual noise components of observed series. The class of priors also naturally incorporates uncertainty about model order, and hence leads in posterior analysis to model order assessment and resulting posterior and predictive inferences that incorporate full uncertainties about model order as well as model parameters. Analysis also formally incorporates uncertainty, and leads to inferences about, unknown initial values of the time series, as it does for predictions of future values. Posterior analysis involves easily implemented iterative simulation methods, developed and described here.

One motivating applied field is climatology, where the evaluation of latent structure, especially quasi-periodic structure, is of critical importance in connection with issues of global climatic variability. We explore analysis of data from the Southern Oscillation Index (SOI), one of several series that has been central in recent high-profile debates in the atmospheric sciences about recent apparent trends in climatic indicators.

Key words: Autoregressive component models, Bayesian time series, Dynamic linear models, Latent time series structure, Quasi-periodic time series components, Time series decompositions

1 Introduction

Recent developments in time series decomposition have focused modelling attention on latent component processes underlying observed time series. This leads naturally to the development of novel approaches to prior modelling in which prior distributions are specified directly on parameters defining such latent component subseries. This approach to prior specification is introduced and developed here, together with resulting developments of posterior inference, model assessment and forecasting, in the context of linear autoregressions.

In recent years Bayesian work in time series has been substantially impacted by computational methods, especially Markov chain Monte Carlo (MCMC) methods, as have all areas of statistical

modelling. In autoregressive models, for example, this is illustrated by MCMC schemes for posterior and predictive inference developed by Albert and Chib (1993), and McCulloch and Tsay (1993, 1994). Such developments apply standard MCMC methods to AR models using prior distributions that are essentially standard normal or uniforms on AR coefficients, though sometimes with imposed constraints to stationary regions. More recently, Barnett, Kohn and Sheather (1996a,b) have developed novel prior models in AR and ARMA models with a primary focus on priors specified for partial regression coefficients rather than raw autoregressive parameters. This is partly motivated by uncertainty about model order, with prior distributions that permit partial regression coefficients at arbitrary lags to be zero. This is an example of how the ability to analyse mathematically complicated prior and posterior distributions via MCMC methods provides freedom and flexibility in developing structured prior distributions in practically meaningful and interpretable parametrisations. We take this position here, but develop alternative classes of priors focussed on the underlying latent component structure of autoregressions and defensible quantitative forms of prior information about the form and nature of such components.

We begin by discussing the implications of structural decompositions of time series introduced in West (1997a) and investigate their relevance in focusing prior modelling attention on latent components and the characteristic root structure of AR processes. This leads into development of classes of prior distributions over parameters that are primary descriptors of the latent structure. These priors have the following characteristics and implications:

- They permit arbitrary collections of real and complex conjugate pairs of characteristic roots, as well as arbitrary values of the moduli and frequencies of such roots. Experience and prior knowledge in an application area will usually lead to expectations about anticipated patterns, such as the likelihood of sustained low frequency trends, quasi-periodic components and the relative levels of high-frequency noise contributions, that can be incorporated in these priors on component structures.

- They allow zero values among the characteristic roots, so catering for prior uncertainty about model order up to a specified maximum value. Here an important feature is that our structured priors naturally allow for the possibility of models of very high order, such as in approximation to other ARMA processes. Indeed, with n observations we can fit and assess AR(p) models of order up to and including $p = n$. In this sense, we are in a framework of almost ideal non-parametric linear time series modelling.

- They allow unit roots, and so cater for persistent low frequency trends and sustained quasi-periodic components in the observed process.

- They incorporate unobserved initial values of the data process as uncertain latent variables, so that all resulting inferences are formally based on incorporating full uncertainties about initial values.

- They induce corresponding priors over the standard autoregressive parameters, concentrated on the stationary region and with interesting and mathematically complicated forms. Though very much of secondary interest in their own right, these implied priors on AR coefficients can generate useful insights and are easily explored by direct prior simulation.

- A specific prior in the class is identified by a small number of hyper-parameters, which may be chosen based on specific forms of quantitative prior information. Alternatively, and usually, these hyper-parameters can be assigned essentially uniform or "reference" prior distributions themselves, so inducing what may be viewed as a default or benchmark analysis.

It is of interest to make comparisons with priors specified on either the standard linear AR coefficients or the corresponding partial autocorrelation coefficients, the latter as illustrated in parallel work of Barnett, Kohn and Sheather (1996a), for example. In addition to a direct focus on meaningful component structures, our approach addresses issues about unit roots. More standard parametrisations can allow for unit roots but, by comparison, it is more difficult to develop flexible classes of priors over the number of unit roots, and the balance between real and complex components, especially with high order autoregressions. Furthermore, the component focus provides much scope for the future development of more structured priors that incorporate substantive prior information, perhaps based on past data, and hierarchical specifications for components. For example, in analysing patient-specific biomedical time series in which we anticipate quasi-periodic bio-rhythmics components, it will be natural to explore informed priors encoding views about likely ranges of periods based on prior studies of other individuals. Though we do not develop this line of thinking further in the current paper, it is clear that such ideas will be immediately accessible using component priors, but very difficult to encapsulate in priors in the more standard parametrisations. Further discussion of prior specification and structure is given in Section 3.

Following development of these classes of priors, we move to issues of posterior analysis and posterior computations using "customised" MCMC methods. This is followed with an illustration in application to a climatological times series. In addition to addressing a specific scientific issue in climatology, this study illustrates and elaborates on the benefits of the novel class of priors through aspects of prior specification, summary posterior and predictive inferences, and model order assessment and uncertainty analysis.

2 Time series decompositions and latent components

Suppose a time series $\{x_t\}$ of equally spaced observations follows a standard autoregression of order p, or AR(p), i.e.,

$$x_t = \Sigma_{i=1}^{p} \phi_i x_{t-i} + \epsilon_t$$

where $\{\epsilon_t\}$ is a sequence of zero-mean uncorrelated innovations and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$ is the constant AR coefficient vector. The innovations are assumed to be normal with constant variance, $\epsilon_t \sim N(\epsilon_t|0, \sigma^2)$. Introduce the characteristic polynomial

$$\phi(u) = 1 - \Sigma_{j=1}^{p} \phi_j u^j = \prod_{i=1}^{p} (1 - \alpha_j u)$$

where $\{\alpha_1, \ldots, \alpha_p\}$ are the the reciprocals of the roots of $\phi(u) = 0$. Assuming a stationary process, $|\alpha_j| < 1$ for each j. We explore stationary models extended to allow arbitrary combinations of unit roots, i.e., $|\alpha_j| = 1$ for some j. When dealing with real data, series may be non-stationary but we can then appeal to the usual methods of differencing or detrending, or, often more appropriately, develop more general models in with the AR(p) process is one component and other describe non-stationary trends and so forth (e.g., West and Harrison 1997, especially chapters 9 and 15). With the backshift operator B, we have the traditional representation $\phi(B)x_t = \epsilon_t$ or

$$\prod_{i=1}^{p} (1 - \alpha_j B) x_t = \epsilon_t.$$

Suppose that the roots α_j are distinct and occur as C pairs of conjugate pairs and $R = p - 2C$ real roots. Write the complex pairs as $r_j \exp(\pm i\omega_j)$ for $j = 1, \ldots, C$, and the real roots as simply r_j for $j = 2C + 1, \ldots, p$; here the r_j and ω_j are real quantities with $\omega_j > 0$.

The basic decomposition result of interest here may be derived theoretically in various ways. As the process is stationary we invert this representation to give

$$x_t = \prod_{i=1}^{p} (1 - \alpha_j B)^{-1} \epsilon_t,$$

which, via partial fractions expansion, becomes

$$x_t = \Sigma_{j=1}^{C} z_{tj} + \Sigma_{j=2C+1}^{p} a_{tj}$$

where the z_{tj} and a_{tj} are latent processes corresponding to the complex and real roots respectively. Corresponding to real roots $j = 2C+1, \ldots, p$ we have

$$(1 - r_j B) a_{tj} = b_j \epsilon_t$$

for some real constants b_j; thus the a_{tj} are correlated AR(1) processes. Corresponding to complex conjugate pairs of roots $j = 1, \ldots, C$ we have

$$(1 - 2r_j \cos(\omega_j) B + r_j^2 B^2) z_{tj} = (d_j + e_j B) \epsilon_t$$

for further real constants d_j and e_j; thus the z_{tj} are correlated ARMA(2, 1) processes in which the AR component represents quasi-periodic behaviour of characteristic frequency ω_j, i.e., period or wavelength $2\pi/\omega_j$. Note that the latent a_{tj} and z_{tj} are driven by functions of the same innovation ϵ_t, and thus the component-specific innovations are inherently highly dependent. Nevertheless, identified components may often have physically meaningful interpretations when considered in the context of the scientific problem generating the data. This is clearly illustrated in the example of section 5.

The original development of the decomposition through state-space models can be found in West (1997a), and is elaborated in West and Harrison (1997, section 9.5). These works stress the constructive nature of the development in the state-space framework. That is, for a given AR vector ϕ and data values $\{x_t\}$, the values of the latent processes $\{a_{tj}\}$ and $\{z_{tj}\}$ may be easily and directly computed using that approach. This is further illustrated in various applications and extensions to time-varying parameter models in, for example, Prado and West (1997), West and Harrison (1997, chapter 15), and West, Prado and Krystal (1999). The applications in these works utilise standard normal or reference priors. By contrast, our focus here is to exploit the decomposition in exploring and implementing new classes of structured priors, with goals and benefits as outlined in the introduction above.

3 Prior Modelling

3.1 General comments

In practical time series modelling the above decomposition and root structure provides a much more natural framework for prior specification than the usual linear model form. In many applications, AR models are used as purely empirical models capable of flexibly representing observed structure, and this encourages the use of high-order models. Success with vector autoregressions is testament to the utility of strategies based on this view, and the successful use of high-order models is enabled through the use of structured smoothness priors (Litterman 1986a,b). Often, however, underlying time series structure may be adequately captured by more parsimonious models and then high-order

autoregressions imply immediate over-fitting and approximations, but within such structure the underlying and physically meaningful patterns will emerge. One example is that of approximation a low-order ARMA model via a high-order AR model. In such cases, the autoregressive roots and resulting component structure of the ARMA process will pass through to the root and component structure of the high-order AR model. The additional root/component structure in the latter represents the residual AR approximation to the MA component of the process. This relates closely to a second key example, in which a low-order AR model is observed with additive noise, very much the norm rather than the exception in reality. Suppose, for example, the data series is generated as $x_t = y_t + \nu_t$ where y_t is AR(q) with roots $\alpha_j, j = 1, \ldots, q$, and ν_t is a zero-mean, uncorrelated sequence. The implied model for x_t is ARMA($q, q-1$) with the same AR root structure, and can be well-approximated by a high-order AR(p) approximation in which these roots are supplemented by others, $\alpha_{q+1}, \ldots, \alpha_p$. Further, the additional roots will have generally small moduli and, if complex, large arguments, reflecting the fact that they describe high-frequency noise characteristics induced by the ν_t. The resulting decomposition of observed time series under the high-order model will deliver estimated latent components corresponding to the decomposition of the "real" AR process y_t, and additional components corresponding to these "noise" contributions. In many practical contexts, especially in the natural and engineering sciences, a small number of physically meaningful components – persistent low frequency and interpretable quasi-periodic components – are realised and can be identified this way.

This discussion motivates the view that, prior to future studies, fitting higher-order AR models be performed in the context of prior beliefs predicated on a smaller number of latent AR(1) and ARMA(2,1) processes being really relevant and perhaps expected, with ranges of amplitudes, moduli and arguments, and that additional components induced by the higher order model will tend to exhibit lower moduli and higher arguments, and often lower amplitudes, consistent with the above interpretations in the context of measurement errors and model approximations. The prior modelling framework in the next section builds on this reasoning.

Before proceeding to develop the prior on the root structure, note that we will be working with a marginal prior on the innovation variance σ^2 that will be taken as either an inverse gamma or a reference prior. Throughout suppose that such a prior, with density $p(\sigma^2)$, is specified, and that σ^2 is independent of the root parameters and initial values.

3.2 Priors on autoregressive root structure

Begin by assuming and specifying fixed upper bounds C and R on the numbers of complex root pairs and real roots, respectively. This implies a value $p = 2C + R$ as an upper bound on model order. Note C and R, hence p, may be quite large. Conditional on C and R, we begin by assuming independent priors for the component roots α_j, distinguishing between real and complex cases. The specific prior class here assume real roots to be exchangeable, and complex roots to be exchangeable. The relevant marginal priors for each case are as follows.

(a) Real Roots

For each $j = 1, \ldots, R$, the real root r_j has prior over support $|r_j| \leq 1$ with density

$$r_j \sim \pi_{r,-1} I_{(-1)}(r_j) + \pi_{r0} I_0(r_j) + \pi_{r1} I_1(r_j) + (1 - \pi_{r0} - \pi_{r,-1} - \pi_{r1}) g_r(r_j)$$

where $I_{\cdot}(\cdot)$ is the indicator function, e.g., $I_0(r) = 1$ if $r = 0$ and 0 otherwise, and $g_r(\cdot)$ is a continuous density over $(-1, 1)$. Here the $\pi_{r,\cdot}$ are prior probabilities at $r_j = -1, 0$ and 1 respectively. Thus the marginal prior for r_j allows for a root on the stationary boundary $r_j = \pm 1$. The prior point mass at

$r_j = 0$ puts non-zero prior probability on a zero root, so that the number of real roots may reduce below the specified maximum R. This partially caters for model order uncertainty, and leads to a posterior distribution for the number of non-zero real roots as a result. The continuous part of the mixture prior, $g_r(\cdot)$, specifies the conditional prior on the full stationary region $-1 < r_j < 1$. The natural default or "reference" prior is the uniform, $g_r(\cdot) = U(\cdot|-1, 1)$, i.e., the formal reference prior for the component AR(1) coefficient r_j truncated to the stationary region. Though the framework is quite general and other priors may be used, we adopt the uniform as the operational default prior and use it in analysis below.

The roots are not identified in the mathematical sense. Under the specified prior and with independence of the r_j, the model is unchanged under arbitrary permutations of the root index j. This could be addressed by modifying the prior so that the non-zero real roots are ordered – i.e., taking the roots to be the order statistics of a random sample from the marginal distribution above – which provides identification. In practice, this can be effectively imposed by exploring posterior inferences about the *ordered* r_j under the exchangeable prior. As posterior computations are via simulation, this is trivially done by simply ordering the sets of simulated roots in the analysis. Notice that very similar ideas are used to trivially resolve related identification problems in mixture modelling (West 1997b). Comments on how this ordering issue relates to the convergence of our MCMC methods is deferred to the final paragraph of section 4 below.

For fixed values of the key probabilities $\pi_{r,\cdot}$ we have implied priors such as that on the number of non-zero roots; that is simply binomial $Bn(R, 1 - \pi_{r,0})$. By considering this, we may be led to evaluate suitable values of $\pi_{r,0}$. Similar comments apply to the other two point masses. However, in most applications we will prefer to treat the $\pi_{r,\cdot}$ as hyperparameters to be estimated, and to impose priors on these probabilities that are relatively or absolutely uniform and viewed as uninformative. For example, a natural choice is the uniform Dirichlet distribution for the three probabilities together,

$$Dir(\pi_{r0}, \pi_{r1}, \pi_{r,-1}|1, 1, 1),$$

and this is used in the example below. In addition to defining a diffuse reference prior on these hyperparameters, it turns out to simplify some of the posterior simulations and computations.

(b) Complex Roots

The prior structure for each pair of complex roots is specified with the same qualitative features as that for the real roots. For each $j = 1, \ldots, C$, we take the defining parameters (r_j, λ_j) where r_j is the modulus and $\lambda_j = 2\pi/\omega_j$ is the period of the quasi-periodic component z_{tj} corresponding to this root pair. The prior is specified over support $0 \leq r_j \leq 1$ and $2 < \lambda_j < \lambda_u$ for a specific upper bound λ_u on periods, and with r_j and λ_j as conditionally independent. Each λ_j is bounded below by 2 as ω_j must be less that π, and λ_u is assumed fixed and by default set to $n/2$, the maximum observable period in a time series of length n. For the modulus, we assume a prior

$$r_j \sim \pi_{c0} I_0(r_j) + \pi_{c1} I_1(r_j) + (1 - \pi_{c1} - \pi_{c0}) g_c(r_j)$$

with components as follows. The first component corresponds to a zero root, with prior probability π_{c0}, so allowing the number of complex components to be fewer than the specified C. Coupled with the prior on real roots, we now have a full prior expression of uncertainty about model order. The second component of the prior for r_j here has non-zero prior probability π_{c1} on the stationary boundary. The final component $g_c(r_j)$ is a continuous distribution on $0 < r_j < 1$, the conditional prior on the stationary region. The first stage prior is completed by specifying a marginal density $h(\lambda_j)$ for the period of the complex root on the specified support $(2, \lambda_u)$. In ranges

of empirical studies with various data sets, we have experimented with various choices of $g_c(r_j)$ and $h(\lambda_j)$, including uniform priors and margins for λ_j based on uniform priors for the corresponding frequency $\omega_j = 2\pi/\lambda_j$. The natural default prior is the "component reference prior" induced by assuming a uniform prior for the implied AR(2) coefficients $2r_j \cos(2\pi/\lambda_j)$ and $-r_j^2$, but restricted to the finite support of λ_j for propriety. This is defined by $g_c(r_j) \propto r_j^2$, so that the margin for r_j is $Beta(\cdot|3,1)$, and $h(\lambda_j) \propto \sin(2\pi/\lambda_j)/\lambda_j^2$ on $2 < \lambda_j < \lambda_u$. In fact this form of $h(\cdot)$ is quite similar to that induced by a proper uniform on frequency ω_j. We take these component reference priors as the operational default priors and illustrate their use in analysis below. Further discussion of prior distributions follows in the next subsection.

The complex component prior specification is completed with structure for the hyperparameters. As with the real roots, we specify a uniform Dirichlet hyperprior on the selection probabilities i.e.,

$$(\pi_{c0}, \pi_{c1}) \sim Dir(\pi_{c0}, \pi_{c1}|1,1).$$

Note also that, again as with the real roots, the complex roots are not identified. Under this prior and assuming independence across j as we do, the model is unchanged under arbitrary permutations of the root index j. In posterior analysis, identification is again imposed by ordering the roots in terms of increasing order of either moduli or period. This is further discussed in the applied study in Section 5.

3.3 Aspects of implied prior distributions

As mentioned above, the priors specified on the root structure induce priors on the numbers of real and complex roots, and so on model order up to the specified maximum. Under the uniform Dirichlet priors for the sets of probabilities this is particularly simple. Write C_+ for the number of non-zero complex root pairs, and R_+ for the number of non-zero real roots. Then the uniform priors on π_{c0} and π_{r0} imply discrete uniform priors on C_+ (over $0, 1, \ldots, C$), and R_+ (over $0, 1, \ldots, R$). Consider now model order $p_+ = 2C_+ + R_+$. Then if $C = 0$ or $R \leq 1$, the prior for p_+ is a discrete uniform on $\{0, 1, \ldots, 2C + R\}$. In general, p_+ does not have uniform prior due to the fact that complex roots occur as pairs. For example, if $C = R = 2$, then $Pr(p_+) = 1/9$ for $j = 0, 1, 3, 5, 6$ but $Pr(p_+) = 2/9$ for $j = 2, 4$.

The specified priors also induce priors, of complicated mathematical forms, on the standard linear autoregressive parameters $\boldsymbol{\phi}$. Consider first the key special cases of strictly stationary AR(1) and AR(2) models. By construction, the continuous components of the corresponding priors, $g_r(\cdot)$ and $(g_c(\cdot), h(\cdot))$, induce priors on the standard AR coefficients that are uniform over, and restricted to, the stationary regions. In this sense, the priors are naturally diffuse in both the root and AR coefficient parametrisations for these two basic models, while constraining to stationarity. The mixtures with point masses on the boundaries of the stationary regions naturally extend these reference priors.

Consider now a more elaborate AR(4) model with exactly one quasi-periodic component, so that $C = 1$ and $R = 2$. For graphic illustration take

$$\pi_{r0} = \pi_{r1} = \pi_{r,-1} = \pi_{c,0} = \pi_{c,1} = 0$$

so that the prior constrains to strictly stationary models. We explore the implied prior on the four implied AR coefficients $\boldsymbol{\phi}$ by simulation: given a random draw from prior for the roots, we can trivially compute the corresponding value of $\boldsymbol{\phi}$. Figures 1 and 2 displays all one- and two-dimensional margins of a sample of 5,000 draws from the prior. Note that we do not need to

impose an identifying ordering on the roots as the AR vector $\boldsymbol{\phi}$ is the same irrespective of ordering. By construction, the prior for $\boldsymbol{\phi}$ is naturally constrained to the stationary region and so the shapes in Figure 1 are contained in this region. Note also that, though based on component priors that are reference for the root parameters, the induced prior on $\boldsymbol{\phi}$ is naturally not uniform.

A final comment relates to one other interesting implication of the conditional prior distribution $h(\cdot)$ for periods λ_j of the quasi-periodic components. This component reference prior density $h(\lambda_j) \propto \sin(2\pi/\lambda_j)/\lambda_j^2$ is unimodal with a mode near $\lambda_j = 2.7$, decreasing smoothly, and slowly, thereafter. Ordering periods for identification implies that any set of *identified* λ_j are initially distributed as the order statistics of a random sample from this prior. As a result, the prior qualitatively implies that any set of such periods will tend to have a few larger values but with more small ones clustered in the 2-4 ranges. This is exactly the pattern anticipated, as discussed in Section 3.1, based on applications of higher order AR models in the past; a few larger periods corresponding to physically meaningful structure, together with several or many high-frequency components representing noise contributions or model approximation. Hence the component reference prior is attractive from this empirical and modelling viewpoint, as well as having the theoretical reference prior justification.

4 Posterior Structure and Analysis

Under the prior structure just described, posterior and predictive calculations are available via Markov Chain Monte Carlo (MCMC) methods (e.g., Gilks et al 1996). The structure of relevant conditional posterior distributions is briefly outlined here, with some commentary on details of our chosen MCMC methods. Further details appear in the Appendix.

First some notation. Write $\mathbf{X} = \{x_1, \ldots, x_n\}$ for the observed time series and, given the maximum model order p, write $\mathbf{Y} = \{x_0, x_{-1}, \ldots, x_{-(p-1)}\}$ for the latent initial values. Our analysis includes formal inference on these initial values. The model parameters are denoted by

$$\boldsymbol{\psi} = \{\alpha_j,\ j = 1, \ldots, p;\ (\pi_{r,-1}, \pi_{r0}, \pi_{r1});\ (\pi_{c0}, \pi_{c1});\ \sigma^2\}.$$

Posterior inferences are based on summarising the full posterior $p(\boldsymbol{\psi}, \mathbf{Y}|\mathbf{X})$. For any subset $\boldsymbol{\xi}$ of elements of $\boldsymbol{\psi}$, let $\boldsymbol{\psi}\backslash\boldsymbol{\xi}$ denote the complementary elements, i.e., $\boldsymbol{\psi}$ with $\boldsymbol{\xi}$ removed. Our MCMC method is based on a standard Gibbs sampling format in which we iteratively simulate elements of $\boldsymbol{\psi}$ and $\mathbf{Y}$ from their conditional posteriors with all conditioning parameters fixed at their latest sampled values. Specifically,

- for each $j = 2C+1, \ldots, r$, sample the real roots individually from $p(r_j|\boldsymbol{\psi}\backslash r_j,\ \mathbf{X}, \mathbf{Y})$;
- for each $j = 1, \ldots, C$, sample the complex roots individually from $p(r_j, \lambda_j|\boldsymbol{\psi}\backslash(r_j, \lambda_j),\ \mathbf{X}, \mathbf{Y})$;
- sample the hyperparameters from conditionally independent posteriors

  $$p(\pi_{r,-1}, \pi_{r0}, \pi_{r1}|\boldsymbol{\psi}\backslash(\pi_{r,-1}, \pi_{r0}, \pi_{r1}),\ \mathbf{X}, \mathbf{Y})$$

  and

  $$p(\pi_{c0}, \pi_{c1}|\boldsymbol{\psi}\backslash(\pi_{c0}, \pi_{c1}),\ \mathbf{X}, \mathbf{Y});$$

- sample the innovation variance from $p(\sigma^2|\boldsymbol{\psi}, \mathbf{X}, \mathbf{Y})$;
- sample the latent initial values from $p(\mathbf{Y}|\boldsymbol{\psi}, \mathbf{X})$.

Each of these distributions is now briefly described.

(1) Conditionals for real roots

Consider any real root $\alpha_j = r_j$, for some j between $2C+1$ and p. Given $\boldsymbol{\psi}\backslash\alpha_j$ and $\mathbf{Y}$, we can directly compute the filtered time series $u_{jt} = \prod_{i=1, i\neq j}^{p}(1-\alpha_i B)x_t$ for $t = 0, \ldots, n$. Under the model it follows that u_{jt} is an AR(1) process with AR parameter r_j and innovation variance σ^2, and with known initial value. The conditional likelihood function for r_j has the standard form, providing a normal kernel in r_j. Under the mixture prior, this leads, as further detailed in Appendix A.1, to the mixture posterior

$$p_{j0}I_0(r_j) + p_{j,-1}I_{(-1)}(r_j) + p_{j1}I_1(r_j) + (1 - p_{j0} - p_{j,-1} - p_{j,1})N_t(r_j|m_j, M_j)$$

where $N_t(\cdot|m, M)$ denotes the density of the normal $N(\cdot|m, M)$ truncated to $(-1, 1)$, and the values (m_j, M_j) and point masses $p_{j0}, p_{j,-1}$ and p_{j1} are trivially computed. This mixture posterior is easily sampled, including direct simulation of the truncated normal by c.d.f. inversion.

(2) Conditionals for complex roots

Index $j = 1, \ldots, C$ identifies the pair of complex conjugate roots $(\alpha_{2,j-1}, \alpha_{2j})$ with parameters (r_j, λ_j). Let A_j be the index set of all other roots, $\boldsymbol{\psi}\backslash(r_j, \lambda_j)$. Then, given $\boldsymbol{\psi}\backslash(r_j, \lambda_j)$ and $\mathbf{Y}$, we can directly compute the filtered time series $w_{jt} = \prod_{i\in A_j}^{p}(1 - \alpha_i B)x_t$ for $t = -1, 0, \ldots, n$. Under the model it follows that w_{jt} is an AR(2) process with AR parameters $\phi_{j1} = 2r_j \cos(2\pi/\lambda_j)$ and $\phi_{j2} = -r_j^2$, innovation variance σ^2, and with known initial values. The conditional likelihood function for (r_j, λ_j) is easily computed at any point, and has a standard form in (ϕ_{j1}, ϕ_{j2}) providing a bivariate normal kernel in these parameters. Sampling the resulting conditional posterior directly is difficult, and here we use a reversible jump Markov chain Monte Carlo step nested within the overall Gibbs iterations. This method is able to jump between different subspaces of the parameter space corresponding to different components in a mixture of continuous distributions with point masses, and so is most suitable for exploring the conditional posteriors here. Technical details appear in Appendix A.2.

(3) Conditionals for hyperparameters

Given sampled values of all the roots, the independent Dirichlet priors for $(\pi_{c0},\ \pi_{c1})$ and $(\pi_{r,-1}, \pi_{r0}, \pi_{r1})$ update to conditionally independent Dirichlet posteriors. Let (r_{-1}, r_0, r_1) be the number of real roots equal to minus one, zero and one, respectively. Then

$$(\pi_{r,-1}, \pi_{r0}, \pi_{r1})|\boldsymbol{\psi}\backslash(\pi_{r,-1}, \pi_{r0}, \pi_{r1}),\ \mathbf{X}, \mathbf{Y} \quad \sim \quad Dir(\cdot|r_{-1}+1, r_0+1, r_1+1).$$

Similarly, and independently,

$$(\pi_{c0}, \pi_{c1})|\boldsymbol{\psi}\backslash(\pi_{c0}, \pi_{c1}),\ \mathbf{X}, \mathbf{Y} \quad \sim \quad Dir(\cdot|c_0+1, c_1+1)$$

where (c_0, c_1) are the numbers of complex pairs of roots with moduli zero and one, respectively.

(4) Conditional for innovation variance

Given the roots α_j, we compute the implied AR parameter vector $\boldsymbol{\phi}$ and hence the imputed innovations $\epsilon_t = x_t - \Sigma_{j=1}^{p}\phi_j x_{t-j}$, making use of the currently imputed initial values. Then the conditional likelihood function for σ^2 has a standard inverse gamma kernel. Computations are

then particularly easy, and standard, if the prior is a conjugate inverse gamma distribution or the standard reference prior.

(5) Conditional for latent initial values

Conditional on $\boldsymbol{\psi}$ and $\mathbf{X}$, we exploit the reversibility of the AR model to deduce the conditional distribution for initial values for their simulation. We sequentially sample $x_0, x_{-1}, \ldots, x_{-(p-1)}$ in turn, conditioning on the most recent sampled values in the reverse-time model $x_t = \Sigma_{j=1}^{p} \phi_j x_{t+j} + \epsilon_t$ for $t = 0, \ldots, -(p-1)$, sampling new innovations ϵ_t at each step. Here, as in (4), the current roots α_j imply values for the ϕ_j. This is essentially the same operation as used in sampling future values x_{n+k} for $k > 0$ in forecasting ahead from the end of the data in the forward-time model. We use this in various forecasting activities, as mentioned below. It is of interest and importance to note that this reverse-time method applies even in cases of unit roots in the AR structure. Even though the AR process is not strictly stationary in such cases, it turns out that the reverse-time model does in fact produce the correct conditional distributions for latent initial values. Appendix A.3 provides discussion and proof of this general and important result.

Finally, note that the model structure and MCMC strategy outlined are such that the resulting Markov chain is trivially seen to be aperiodic and irreducible, and so standard theory (e.g., Tierney 1994) guarantees convergence of resulting MCMC sequences. Our experience with these models in ranges of applications indicates that convergence is generally clean and rapid. Note that we run the MCMC analysis on the unidentified component parameters, and that components are identified by appropriately labelling the posterior sampled values. The real roots are identified by ordering according to the r_j, and complex roots may be identified by ordering according to either the periods λ_j or moduli r_j. As in other models with similar identification features, MCMC strategies that run on the unconstrained, unidentified parameters rather typically converge faster as they do not suffer problems of high correlations between consecutively ordered parameters when near the boundaries induced by the constraints (e.g., West 1997b in hierarchical mixture models).

5 Analysis of the Southern oscillation index

The *El Niño-Southern Oscillation* is a climatological phenomenon that has been of some interest in climate change studies in recent decades. The Southern Oscillation Index (SOI) series graphed in Figure 3 consists of 540 observations on the SOI, computed as the "difference of the departure from the long-term monthly mean sea level pressures" at Tahiti in the South Pacific and Darwin in Northern Australia. The monthly index spans the period from 1950-1995, and is related to sea surface temperatures. The fact that most of the observations in the last part of the series take negative values is related to a recent warming in the tropical Pacific. This kind of pattern was of focus in studies of Trenberth and Hoar (1996), who analysed closely related climatological series. A key question of interest is to determine just how unusual such patterns are, and whether or not they are adequately described by stationary models for the SOI.

Our analysis of this series is based on an AR model with maximum order $p = 40$ ($C = 15$, $R = 10$), $\lambda_u = 270$ and $p(\sigma^2) \propto 1/\sigma^2$. We present posterior inferences based on $11,000$ samples from two different MCMC chains with a burn-in of $5,500$ iterations; these summaries are supported by investigations of convergence diagnostics. Convergence is clean and fast, consistent with our experience in ranges of empirical studies.

The marginal posterior distributions for p_+, C_+ and R_+ appear in Figure 4. The posterior assigns appreciable mass to values of p_+ from 8 to 17 representing a high degree of uncertainty on the model order. Though not graphed, the prior on p_+ is flat over this range, and the likelihood

function indistinguishable from the posterior shown. The mode is at 11 (the order chosen by AIC is 10) but the posterior mass is quite diffuse over higher values. The analysis favours 3-5 complex components, of which one is clearly dominant (see below), and several real roots with quite low moduli representing high frequency structure and noise. Figure 5 exhibits the forecasting performance of the fitted model, providing graphs of simulated "futures" over an horizon of 540 observations, the same span as the original data which are also graphed. The four plots of the SOI with a sample of forecasts shows that the model adequately captures the features in the data (one frame at a time, and ignoring the time axis, the reader might try to guess where the true data ends and the sampled futures begin!). Additional posterior residual analyses confirm model adequacy, especially in terms of lack of structure and residual normality of the AR innovations.

A first set of posterior histograms for the moduli and wavelengths of the two dominant complex roots appear in Figure 6. Here the complex roots are identified through an ordering of the r_j. For the first root, the modulus lies around $0.8 - 0.9$, corresponding to a half-life of about $3 - 6$ months, so that the periodicities are fairly heavily damped. The posterior indicates that the corresponding period likely lies between 4 and 5 years, though with some mass on higher values. The 4-5 year range is consistent with historical analyses of related data and generally accepted views about El Niño periodicities. For the second component, the modulus is around $0.6 - 0.8$ and the posterior for wavelength of this component is multi-modal, with modes near 3 and 6 months. The identification problem is evident here. Beyond the first complex component with an evidently high modulus, there are one or two additional components identified and whose moduli are similar. As a result, the ordering by moduli "mixes" these components, resulting in the multi-modality displayed here. The issue is resolved by imposing the alternative identification constraint – ordering by periods λ_j. The resulting histograms for the two components with largest periods appear in Figure 7. The dominant component remains dominant – whether identified by maximum modulus or period – and now the second component is clearly identified as having a period of about 6 months, corresponding to the second harmonic of an annual seasonal pattern in the series. The posterior for the next component (not graphed here) indicates a period of around 3 months, corresponding to the third harmonic of an annual seasonal pattern, and it was the mixing of this with the semi-annual component that induced the multi-modality in Figure 6.

Figure 8 displays posteriors for the two smallest and two largest real roots, with the posterior probabilities of boundary values $-1, 0$ or 1 indicated. The latter roots are essentially zero, while the former have posteriors centred around -0.6 and -0.4 respectively, indicating high-frequency oscillations. There is no evidence of unit roots, either real or complex, in this posterior analysis (interested readers may see Huerta and West 1999 for an example in which unit roots are critical).

Further posterior summaries involve decomposition of the SOI series into latent constituent sub-series, as discussed in Section 2. Ordering latent quasi-periodic components by wavelengths, we compute the elements of the decomposition for every sampled value of the MCMC chain and average across samples. The resulting posterior means, as functions of time, of the dominant 4-5 year quasi-periodic component and the semi-annual component are graphed in Figure 9. Here we also display the SOI series and the posterior means of the AR innovations. All components are graphed on the same scale as the data for comparison, and subsidiary components (not graphed) have relatively very small amplitudes over the time course of the data, validating the claim that the two quasi-periodic components identified are truly "key".

A final study explores predictive consistency of the model with the recent observed behaviour of the SOI. We take a lead from Trenberth and Hoar (1996) who studied (among various other issues) questions about the atypicality of negative "excursions" of the Darwin sea level pressure measurements (SLP), a time series related to the SOI, in recent years. One particular period of

focus is the 5.5 year from 1990-mid 1995, during which the seasonal SLP exhibits a run of negative values that these authors judge is very unlikely to be accounted for solely by natural variability. That analysis uses data going back further than 1950, so our study is not directly comparable. Nevertheless, it is of interest to explore such issues with the SOI series in connection with recent "unusual" climatic patterns. Noting that the 66 month period from 1990-mid 1995 exhibits 50 months of negative values, we first refit the model on data only up to the start of 1990, and then drew repeat "sampled futures" from our analysis. Identifying negative values in such simulations provides a profile of the posterior predictive distribution for negative values in 5.5 year periods under the assumption of stationarity, but averaged over all possible models and so fully accounting for model uncertainty in predictions. This distribution appears in Figure 10, and indicates appreciable support for larger numbers – the chance of 50 or more is about 0.072 based on this simulation. We conclude that the preponderance of negative monthly SOI values in the most recent 5.5 years is a little unusual under our model, but not so extreme as to cast serious doubt on the assumption that the data generating process is stationary – though there is a great deal of posterior uncertainty about model order, the model averaging maintains global stationarity. Relative to the SOI history over 1950-1990, this particular aspect of the recent behaviour is not inconsistent with the natural variability in a stationary linear model context.

6 Closing Comments

This paper considered development, specification and analysis of autoregressive time series models under new classes of prior distributions over parameters characterising the underlying latent structure of such models. The approach permits fully Bayesian inference on autoregressive model order, the number of complex and real roots of an autoregressive model, unit roots, uncertain initial values and parameters of quasi-cyclical latent component processes, in particular. Model analysis is illustrated in a climatological example, where taking into account model order uncertainty leads to predictive inferences that differ markedly from results of previous work based on specific, selected models. Our current research with these models is exploring applied studies to evaluate the framework in various forecasting scenarios as well as in time series decomposition, and in studies of comparisons with more standard AR-ARMA estimation and model selection methods. Additional developments underway include extensions to allow for unequally-spaced time series and continuous time autoregressions.

Acknowledgements

The authors are grateful for useful discussions with Merlise Clyde and Giovanni Parmigiani of ISDS on early versions of this work, and for valuable comments and suggestions of the editor and two referees. This research was partially supported by the National Science Foundation under grants DMS-9704432 and DMS-9707914.

References

Albert, J.H., and Chib, S. (1993) Bayesian inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *J. Bus. and Econ. Statist.,* **11**, 1-15.

Barnett, G., Kohn, R. and Sheather, S. (1996a) Bayesian estimation of an autoregressive model using Markov chain Monte Carlo. *J. Econometrics* **74**, 237-254.

Barnett, G., Kohn, R. and Sheather, S. (1996b) Robust Bayesian estimation of autoregressive-moving average models. *Working Paper,* Statistics Group: Australian Graduate School of Management, University of New South Wales, Australia.

Gilks, W.R., Richardson, S., and Spiegelhalter D.J. (1996) *Markov Chain Monte Carlo in Practice*, Chapman and Hall: London.

Green, P. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika,* **82**, 711-732.

Huerta, G., and West, M. (1999) Bayesian inference on periodicities and component spectral structure in time series. *J. Time Series Anal.,* (to appear).

Litterman, Robert B. (1986a) Specifying vector autoregressions for macro-economic forecasting. In *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti,* P.K. Goel and A. Zellner, eds. North-Holland, Amsterdam.

Litterman, Robert B. (1986b) Forecasting with Bayesian vector autoregressions: Five years of experience. *J. Bus. and Econ. Statist.,* **4**, 25-38.

McCulloch, R., and Tsay, R. (1993) Bayesian inference and prediction for mean and variance shifts in autoregressive time series. *J. Amer. Statist. Assoc.,* **88**, 968-978.

McCulloch, R., and Tsay, R. (1994) Bayesian analysis of autoregressive time series via the Gibbs sampler. *J. Time Series Anal.,* **15**, 235-250.

Prado, R., and West, M. (1997) Exploratory modelling of multiple non-stationary time series: Latent process structure and decompositions. In *Modelling Longitudinal and Spatially Correlated Data,* (ed: T. Gregoire), Springer-Verlag, pp349-362.

Tierney, L. (1994) Markov chains for exploring posterior distributions. *Ann. Statist.,* **22**, 1701-1762.

Trenberth, K.E., and Hoar, T.J. (1996) The 1990-1995 El Niño-Southern oscillation event: Longest on record. *Geophys. Res. Letters,* **23**, 57-60.

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

West, M. (1997a) Time series decomposition. *Biometrika,* **84**, 489-494.

West, M. (1997b) Hierarchical mixture models in neurological transmission analysis. *Journal of the American Statistical Association,* **92**, 587-606.

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

Appendix

A.1. Posterior point masses for AR(1) components

Recall that u_{jt}, as defined in Section 4, is an AR(1) process with AR parameter r_j and innovation variance σ^2. The parameters of the mixture posterior are:

- For the truncated normal component,

$$m_j = (\Sigma_{t=1}^n u_{tj} u_{t-1,j})/\Sigma_{t=1}^n {u_{t-1,j}}^2$$

and

$$M_j = \sigma^2/\Sigma_{t=1}^n {u_{t-1,j}}^2.$$

- For the point masses at -1, 0 and 1, we have

$$p_{j,-1} \propto k_{j,-1} \equiv \pi_{r,-1} \exp(-(-1-m_j)^2)/2M_j),$$

$$p_{j0} \propto k_{j0} \equiv \pi_{r0} \exp(-{m_j}^2)/2M_j)$$

and

$$p_{j1} \propto k_{j1} \equiv \pi_{r1} \exp(-(1-m_j)^2)/2M_j).$$

The marginalisation constant is equal to

$$k_{j0} + k_{j1} + k_{j,-1} + ((1-\pi_{r0}-\pi_{r1}-\pi_{r,-1})/2)\sqrt{2\pi M_j}(\Phi((1-m_j)/\sqrt{M_j}) - \Phi((-1-m_j)/\sqrt{M_j}))$$

where $\Phi(.)$ denotes the standard normal distribution function.

A.2. MCMC methods for AR(2) component parameters

Recall that w_{jt} is an AR(2) process with parameters $\phi_{j1} = 2r_j \cos(2\pi/\lambda_j)$ and $\phi_{j2} = -r_j^2$. Define $\mathbf{D_j}$ to be the matrix with columns $\mathbf{d}_{j1}$ and $\mathbf{d}_{j2}$ where $\mathbf{d}'_{j1} = (w_{j0}, w_{j1}, \ldots, w_{j,n-1})$ and $\mathbf{d}'_{j2} = (w_{j,-1}, w_{j0}, \ldots, w_{j,n-2})$, and then let $\mathbf{H_j} = (\mathbf{D_j}'\mathbf{D_j})^{-1}$ with entries H_{jlk}. Further, define $\mathbf{h}_j = \mathbf{H_j}\mathbf{D_j}'\mathbf{w}_j$ with $\mathbf{w}'_j = (w_{j1}, w_{j2}, \ldots, w_{jn})$ with elements $\mathbf{h}_j = (h_{j1}, h_{j2})'$.

A reversible jump Markov chain Monte Carlo method (e.g., Green 1995) is used to simulate samples from the target conditional posterior distribution. A description of the changes in dimensionality and generation of proposal points follows.

- Suppose the current parameter value is at the origin $\phi_{j1} = \phi_{j2} = 0$. Moves from this state are as follows, each with probability 1/3:
  - Remain at the origin.
  - Jump to new values $\phi_{j2} = -1$ and $\phi_{j1} \sim N(\cdot|m_j, M_j)$ truncated to $(-2, 2\cos(2\pi/\lambda_u))$, where

$$m_j = \Sigma_{t=1}^n((w_{jt} + w_{j,t-2})w_{j,t-1})/(\Sigma_{t=1}^n w_{j,t-1}^2) \quad \text{and} \quad M_j = \sigma^2/(\Sigma_{t=1}^n {w_{j,t-1}}^2).$$

- Jump to new values $\phi_{j2} \sim N(\cdot|h_{j2}, \sigma^2 H_{j,22})$ truncated to $(-1, 0)$, and $\phi_{j1} \sim N(\cdot|m_j^*, M_j^*)$ truncated to $(-2(-\phi_{j2})^{1/2}, 2\cos(2\pi/\lambda_u)(-\phi_{j2})^{1/2})$ where

$$m_j^* = h_{j1} + H_{j,12}(\phi_{j2} - h_{j2})/H_{j,22} \quad \text{and} \quad M_j^* = \sigma^2(H_{j,11} - H_{j,12}^2/H_{j,22}).$$

- If the chain is at a point $(\phi_{j1}, -1)$, the three, equally probable moves are as follows:
  - Jump to the origin $\phi_{j1} = \phi_{j2} = 0$.
  - Make a reverse jump to a point with the same first coordinate ϕ_{j1} but ϕ_{j2} generated from $N(\cdot|h_{j2}, \sigma^2 H_{j,22})$ truncated to $(-1, 0)$.
  - Remain at $\phi_{j2} = -1$ but with a new value $\phi_{j1} \sim N(\cdot|m_j, M_j)$ truncated to the interval $(-2, 2\cos(2\pi/\lambda_u))$ and m_j, M_j as defined above.
- Finally, when the current parameter values are neither at the origin nor on the line with $\phi_{j2} = -1$, the three, equally probable moves are as follows:
  - Jump to the origin $\phi_{j1} = \phi_{j2} = 0$.
  - Jump to $(\phi_{j1}, -1)$ where ϕ_{j1} is fixed at the current value.
  - Jump to new values (ϕ_{j1}, ϕ_{j2}) with $\phi_{j2} \sim N(\cdot|h_{j2}, \sigma^2 H_{j,22})$ truncated to $(-1, 0)$ and $\phi_{j1} \sim N(\cdot|m_j^*, M_j^*)$ truncated to $(-2(-\phi_{j2})^{1/2}, 2\cos(2\pi/\lambda_u)(-\phi_{j2})^{1/2})$ with m_j^* and M_j^* as defined above.

For each of the possibilities a rejection step is also necessary. The computation of the acceptance-rejection probability is based on ratios of the likelihood function, prior and proposal densities evaluated at the candidate and current parameter values.

A.3. Reversibility of AR models with unit roots

Here we prove that we can simply reverse the time arrow to impute latent initial values of AR models even in cases of unit roots.

To be quite general, suppose the AR(p) model has a unit roots so that $\phi(B) = \mu(B)\psi(B)$ where $\mu(B)$ is the polynomial corresponding to the unit roots and, by definition, $\psi(B)$ is the polynomial of a strictly stationary AR(d) process with $d = p - a$. This AR(d) model is that for the filtered series $y_t = \mu(B)x_t$, in that $\psi(B)y_t = \epsilon_t$ for all t. The y_t process is time reversible, so that $\psi(B^{-1})y_t = \epsilon_t^*$ where $\epsilon_t^* \sim N(\cdot|0, \sigma^2)$ independently. This may also be written as $y_t = \psi(B^{-1})^{-1}\epsilon_t^*$ for all t. Now, based on the observed data $x_1, \ldots, x_n$, we have observed values of y_t for $t > a$, and the reverse-time model for y_t allows us to simulate the latent initial values $y_a, y_{a-1}, \ldots,$ backwards in time. These values provide the required initial values for the x_t process by inverting the identities $y_t = \mu(B)x_t$ for each t, which involves studying $\mu(B)$ in more detail.

In general, suppose there are e unit real roots and f pairs of unit-modulus complex roots with arguments (frequencies) ω_j, $(j = 1, \ldots, f)$. Then $a = e + 2f$ and

$$\mu(B) = (1 - B)^e \prod_{j=1}^{f} (1 - \gamma_j B + B^2)$$

where $\gamma_j = 2\cos(\omega_j)$ for $j = 1, \ldots, f$. This expands as

$$\mu(B) = 1 - m_1 B - m_2 B^2 - \cdots - m_{a-1} B^{a-1} + (-1)^a B^a,$$

where the coefficients $m_1, m_2, \ldots, m_{a-1}$ satisfy certain symmetry constraints. In particular, an easy inductive argument delivers the identity $m_j = (-1)^a m_{a-j}$ for $j > a/2$; this turns out to be critical in delivering the reversibility result, as we now show.

In the identity $y_t = \mu(B)x_t$ with t replaced by $t+a$, we now have

$$y_{t+a} = x_{t+a} - \Sigma_{j=1}^{a-1} m_j x_{t+a-j} + (-1)^a x_t$$

so that

$$x_t = (-1)^a \Sigma_{j=1}^{a-1} m_j x_{t+a-j} + (-1)^{a+1} x_{t+a} + (-1)^a y_{t+a}. \tag{1}$$

As $y_{t+a} = \psi(B^{-1})^{-1}\epsilon^*_{t+a}$ the final term in equation (1) is simply equal to $\psi(B^{-1})^{-1}\tilde{\epsilon}_t$ for some independent error sequence $\tilde{\epsilon}_t \sim N(\cdot|0, \sigma^2)$. Further, the coefficient symmetries $m_j = (-1)^a m_{a-j}$ for $j > a/2$ imply that the first term in equation (1) is equivalent to

$$\Sigma_{j=1}^{a-1} m_j x_{t+j},$$

resulting in

$$x_t = \Sigma_{j=1}^{a-1} m_j x_{t+j} + (-1)^a x_{t+a} + \psi(B^{-1})^{-1}\tilde{\epsilon}_t.$$

Now note, simply, that this has precisely the form

$$\mu(B^{-1})x_t = \psi(B^{-1})^{-1}\tilde{\epsilon}_t$$

or, on inverting $\psi(B^{-1})^{-1}$, simply

$$\phi(B^{-1})x_t = \tilde{\epsilon}_t. \tag{2}$$

Equation (2) is the reverse-time version of the initial AR(p) model for x_t. By taking the route through the initial values of the stationary filtered series y_t obtained via reversibility, we have deduced that the required initial latent values of the x_t series may be written in the form of equation (2). Hence, even though the x_t series is not strictly stationary, the recursively defined conditional distributions of the latent initial values are precisely those obtained from the reverse time calculation. This result may be exploited to trivially impute/sample latent initial values even when there are unit roots.

CAPTIONS

Figure 1: Univariate prior margins for $\phi_1, \ldots, \phi_4$ in AR(4) model with $C = 1$ and $R = 2$.

Figure 2: Bivariate samples from prior for $\boldsymbol{\phi}$ in in AR(4) model with $C = 1$ and $R = 2$.

Figure 3: Southern oscillation index. 540 measurements of the difference of the departure from the long-term monthly mean sea level pressures at Tahiti and Darwin. Observations were taken between 1950-1995.

Figure 4: Posterior distributions for model order p_+ and the numbers of complex and real roots in the AR model for the SOI series.

Figure 5: Southern oscillation index and four "sample futures" for the next 540 months.

Figure 6: Posterior distributions for the moduli and wavelength of the two main complex roots (ordered by moduli).

Figure 7: Posterior distributions for the moduli and wavelength of the two main complex roots (ordered by periods).

Figure 8: Posterior distributions for the two largest and two smallest real roots.

Figure 9: Southern oscillation index series with estimated trajectories of the two key quasi-periodic components and the estimated innovation sequence.

Figure 10: Posterior predictive distribution for at least 50 months of negative SOI values in 5.5 years.

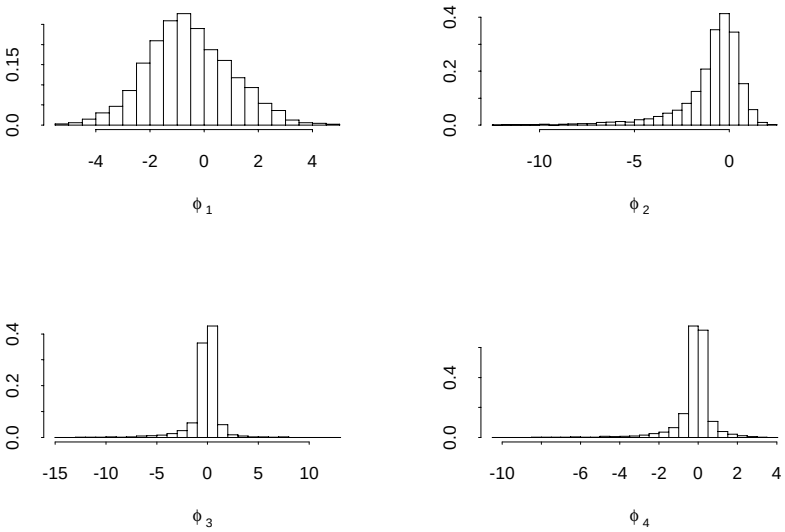

Figure 1:

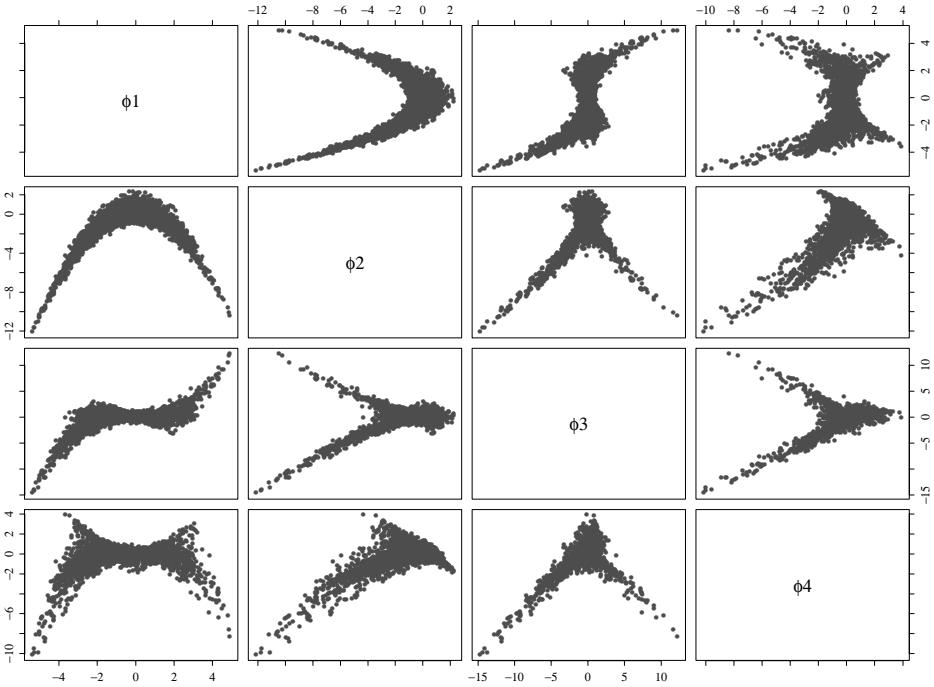

Figure 2:

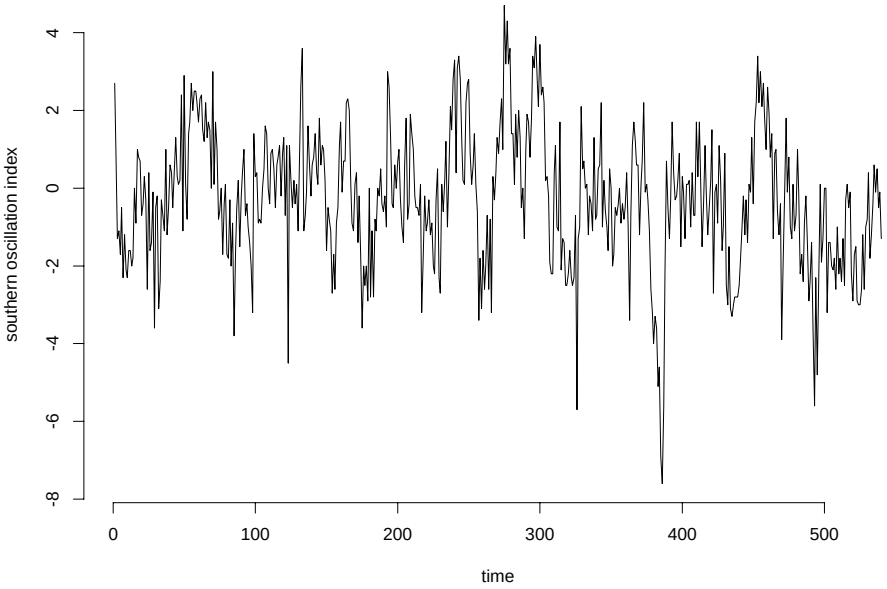

Figure 3:

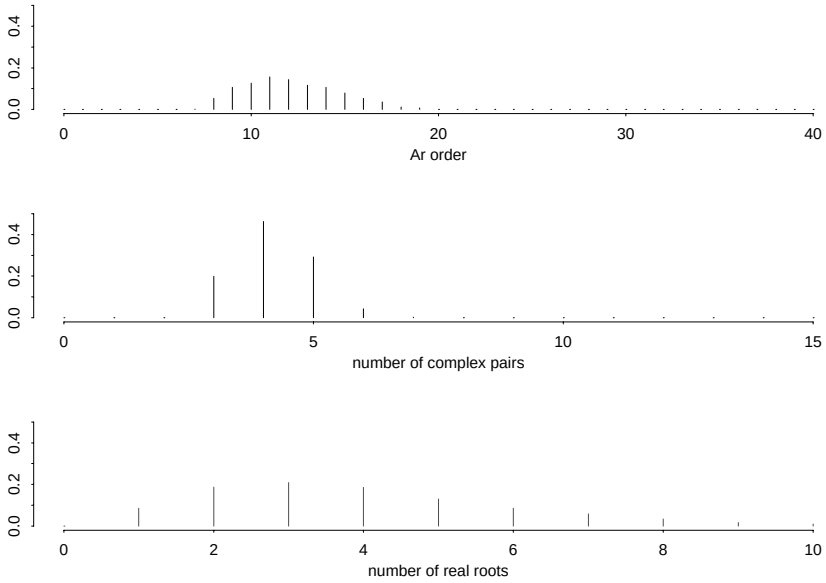

Figure 4:

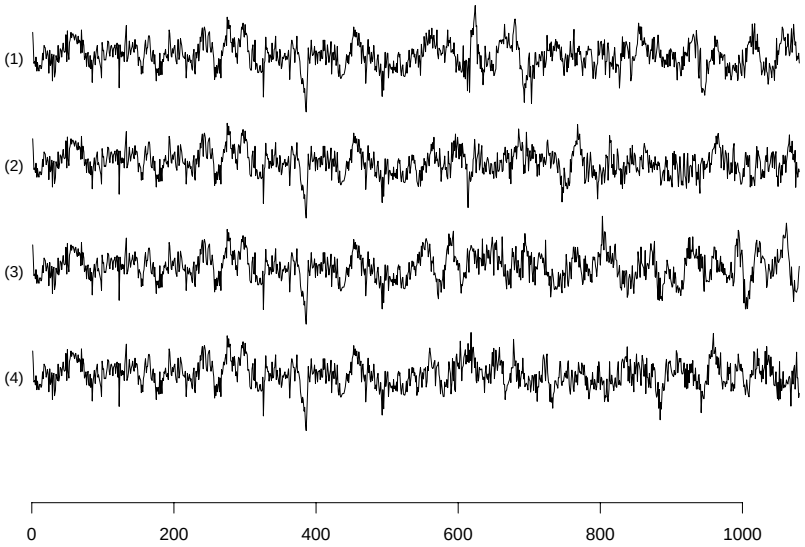

Figure 5:

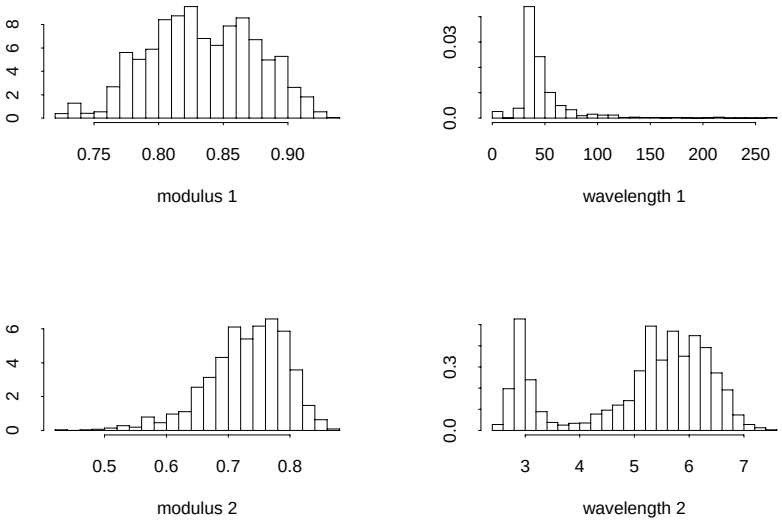

Figure 6:

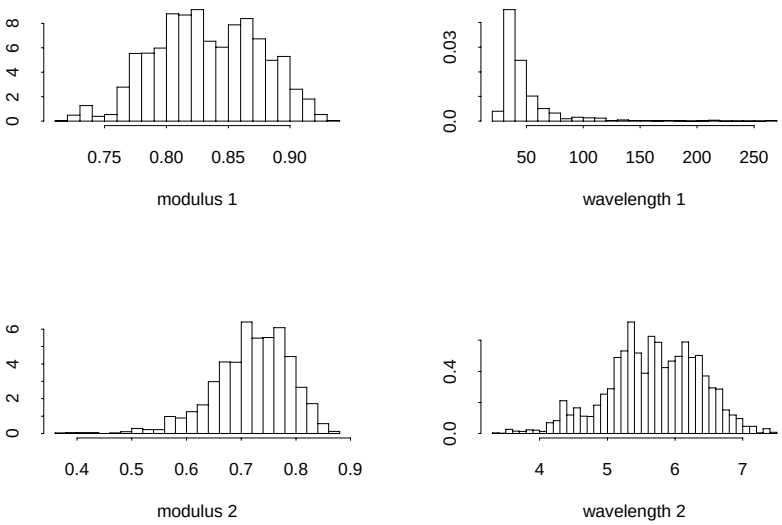

Figure 7:

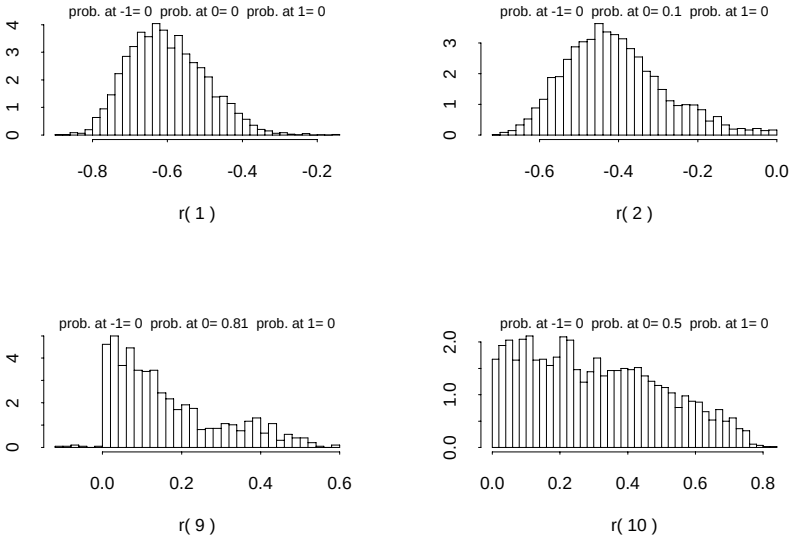

Figure 8:

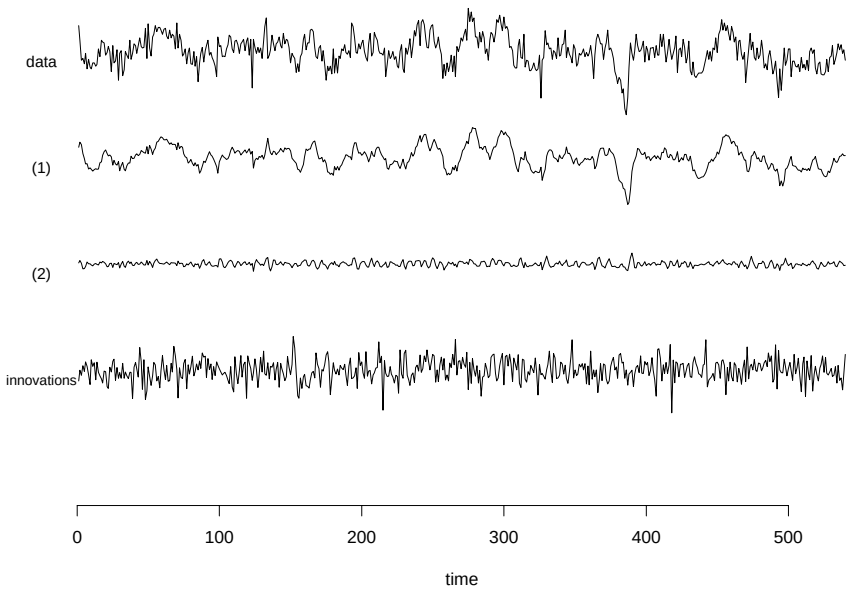

Figure 9:

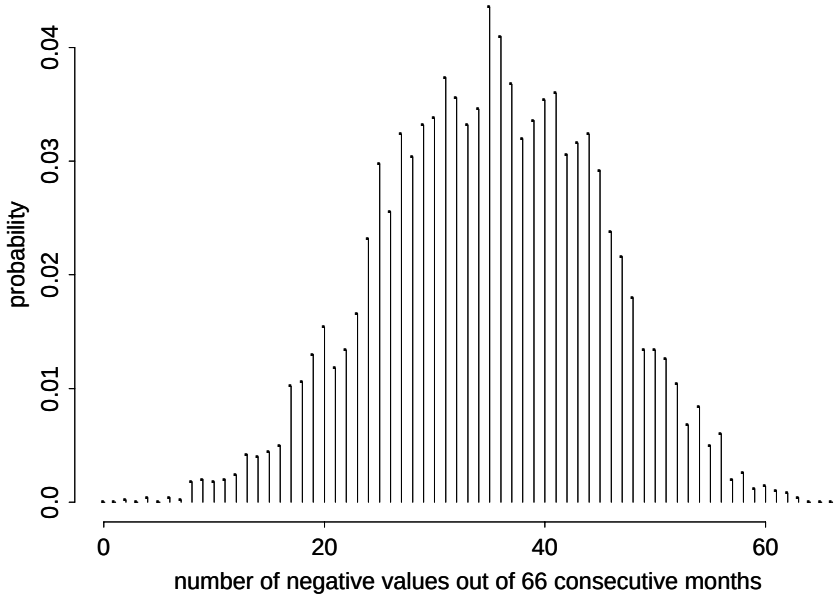

Figure 10: