Bayesian Inference on Periodicities and Component Spectral Structure in Time Series

Gabriel Huerta and Mike West
Institute of Statistics & Decision Sciences,
Duke University, Durham NC 27708-0251
http://www.stat.duke.edu/

Summary

We detail and illustrate time series analysis and spectral inference in autoregressive models with a focus on the underlying latent structure and time series decompositions. A novel class of priors on parameters of latent components leads to a new class of smoothness priors on autoregressive coefficients, provides for formal inference on model order, including very high order models, and leads to the incorporation of uncertainty about model order into summary inferences. The class of prior models also allows for subsets of unit roots, and hence leads to inference on sustained though stochastically time-varying periodicities in time series. Applications to analysis of the frequency composition of time series, in both time and spectral domains, is illustrated in a study of a time series from astronomy. This analyses demonstrates the impact and utility of the new class of priors in addressing model order uncertainty and in allowing for unit root structure. Time domain decomposition of a time series into estimated latent components provides an important alternative view of the component spectral characteristics of a series. In addition, our data analysis illustrates the utility of the smoothness prior and allowance for unit root structure in inference about spectral densities. In particular, the framework overcomes supposed problems in spectral estimation with autoregressive models using more traditional model fitting methods.

Some key words: Autoregressive component models, Quasi-periodic time series, Spectral analysis, Time series decompositions, Unit roots

1 Introduction

Recent work on time series decomposition has stressed the utility of exploring estimated autoregressive models, among others, through inferred latent components of the time series under study. Various related autoregressive component models have been discussed in West (1996, 1997) and West and Harrison (1997, section 9.5). In the current paper, we explore and exploit new classes of smoothness priors, introduced recently in Huerta and West (1997), in developing Bayesian spectral inference. Huerta and West (1997) had an inferential focus on questions about model order and the effects of model order uncertainty on time series decompositions and other inferences. The current paper is, in contrast, concerned with the use of these models and priors in exploring the spectral composition of time series, and spectral density estimation. The basic structure of the model and class of priors is summarised in Section 2, with commentary about motivation, model fitting, and the development of posterior inferences on spectral characteristics. Section 3 discusses inference on spectral characteristics, in both time and frequency domain. Section 4 explores analysis of an astronomical time series using this approach.

2 Time series model class and analysis

2.1 Time Series Models and Decompositions

Begin with a time series $\{x_t\}$ of equally spaced observations modelled via the autoregression of order p, $\phi(B)x_t = \epsilon_t$ where B is the backshift operator, $\phi(u) = 1 - \phi_1 u - \ldots - \phi_p u^p$ is the characteristic polynomial of the model, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$ is the autoregressive coefficient vector, and $\{\epsilon_t\}$ is a sequence of zero-mean, independent and identically distributed innovations with $\epsilon_t \sim N(\epsilon_t|0, \sigma^2)$ for each t. The model may be characterised in terms of the $\boldsymbol{\phi}$ vector or in terms of the reciprocals of the roots of the polynomial $\phi(u) = 0$. Assume these p roots to be distinct and to 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 $j = 2C + 1, \ldots, p$, where the r_j and ω_j are real quantities with $\omega_j > 0$.

As presented in West (1997), we may write

$$x_t = \sum_{j=1}^{C} z_{tj} + \sum_{j=1}^{R} a_{tj}$$

where the z_{tj} and a_{tj} are latent processes corresponding to the complex and real roots respectively. Corresponding to real roots $j = 1, \ldots, R$, the a_{tj} are correlated autoregressive processes of order one, given by $(1 - r_{j'}B)a_{tj} = b_j\epsilon_t$ for some real constants b_j and with $j' = 2C + j$. Corresponding to complex conjugate pairs of roots $j = 1, \ldots, C$, the z_{tj} are correlated autoregressive, moving average processes of order $(2, 1)$, given by

$(1 - 2r_j \cos(\omega_j)B + r_j^2 B^2)z_{tj} = (b_j + d_j B)\epsilon_t$ for further real constants b_j and d_j. The autoregressive component here represents quasi-periodic behaviour at the characteristic frequency ω_j, with characteristic period $\lambda_j = 2\pi/\omega_j$. The state-space representation of autoregressive models provides a simple development of this decomposition and a direct approach to computing the values of the latent processes a_{tj} and z_{tj}, for each j and over all time t, at any specified value of ϕ. Extensions and illustrations appear in Prado and West (1997), and West and Harrison (1997, section 9.5 and chapter 15).

We should stress that the above discussion and decomposition results make no reference to stationarity. Stationarity corresponds to $|r_j| < 1$ for all $j = 1, \ldots, p$, whereas the autoregressive model and above decompositions are valid for *any* values of the r_j, as is clear in the derivation of the decomposition from a state-space viewpoint (West 1997). In the case of a stationary model, the decomposition is exactly that derived from the standard partial fractions inversion of stationary autoregressions, but the state-space approach demonstrates a much broader validity (including, for example, time-varying parameter autoregressions). Here this is important as we are explicitly interested in possible unit roots.

2.2 Prior Modelling

Huerta and West (1997) introduced a class of hierarchical priors on the component structure of autoregressions just described. This framework allows for the fitting of very high order models without running into problems of over-fitting, formal and comprehensive assessment of model order uncertainty, and the assessment of potential unit root structure, among other things. These three specific issues are critical to the use of high order autoregressions in spectral analysis.

The class of priors assumes specified fixed upper bounds C_+ and R_+, hence an upper bound $p_+ = 2C_+ + R_+$ on model order. These may be large. Independent priors are specified on the real roots, the complex roots and the innovations variance. First, each real root r_j has a prior that

- gives probability $\pi_{r,0}$ to $r_j = 0$,
- gives probability $\pi_{r,-1}$ to $r_j = -1$,
- gives probability $\pi_{r,1}$ to $r_j = 1$, and
- is otherwise uniform on the interval -1 to 1, denoted by $U(r_j| -1, 1)$.

A similar prior structure describes the complex conjugate pairs of roots $r_j \exp(\pm i\omega_j)$. Each has a marginal prior that

- gives probability $\pi_{c,0}$ to $r_j = 0$,

- gives probability $\pi_{c,1}$ to $r_j = 1$, and
- otherwise has r_j independent of ω_j and with margins given by
  - $r_j \sim Be(r_j|\beta, 1)$, a beta prior with $\beta \geq 1$, and
  - $\lambda_j \sim U(\lambda_j|2, L)$ where $\lambda_j = 2\pi/\omega_j$ and L is a specified upper bound.

The prior is completed by specifications for the innovation variance σ^2 and the hyperparameters $\pi_{r,0}, \pi_{r,-1}, \pi_{r,1}, \pi_{c,0}, \pi_{c,1}, \beta$ and L. The innovation variance is assumed independent of the root parameters and has a specified marginal prior, usually a conditionally conjugate inverse gamma prior. Priors for the hyperparameters may be specified as context dependent, though specific reference prior forms are suggested for initial analysis, at least. For the point-masses, we favour independent uniform Dirichlet distributions, namely $Dir(\pi_{r,0}, \pi_{r,1}, \pi_{r,-1}|1, 1, 1)$ and $Dir(\pi_{c,0}, \pi_{c,1}|1, 1)$. These priors are used in the example below. In addition to defining diffuse reference priors on these probability hyperparameters, they turn out to simplify some of the posterior simulations and computations. Our reported analyses to date have been based on these prior specifications, and with values of β and L specified, as is illustrated in the application below. The upper bound L on the characteristic periods λ_j is essentially arbitrary. The hyperparameter β is chosen to equal unity so that the marginal prior for the r_j has a uniform density, consistent with the view of a non-informative prior specification.

Note that the prior point masses at $r_j = 0$ allow for the numbers of roots, both complex and real, to fall below the upper bounds specified. This implies that the model order can fall below the upper bound p_+ specified. The application below explores such priors, and the corresponding posteriors. Another noteworthy point is that the roots are not identified. The prior, and the model, is unchanged under arbitrary permutations of the root index j within the set of R_+ real roots, and separately within the set of C_+ complex roots. This is not an issue at the prior modelling stage, but becomes so in posterior analysis as it leads to symmetries and multimodalities in the full posterior distribution. At that point, identification of real roots is imposed simply by relabelling them in order of increasing value. Identification is similarly achieved for the complex roots by relabelling them in a chosen order, either in order of increasing moduli or of increasing frequency, or via some combination of both.

2.3 Posterior and Predictive Analysis

Posterior and predictive calculations are developed using customised Markov chain Monte Carlo methods, which the interested reader can find in Huerta and West (1997). It suffices here to note that efficient iterative simulations methods now exist to implement fully

Bayesian analysis under the model and prior specified. Important features of this analysis include the treatment of latent initial values and missing data. The former is detailed in the above reference, but the development to include missing data is more recent.

Write ψ for all model parameters and hyperparameters, so that

$$\psi = \{\phi;\ (\pi_{r,-1}, \pi_{r0}, \pi_{r1});\ (\pi_{c0}, \pi_{c1});\ \sigma^2\}.$$

Also, write $\mathbf{X}_I = \{x_0, x_{-1}, \ldots, x_{-(p-1)}\}$ for the p latent initial values. Posterior inferences are based on summarising the full posterior $p(\psi, \mathbf{X}_I|\mathbf{X})$ based on the observed time series data $\mathbf{X} = \{x_1, \ldots, x_n\}$. Our MCMC method is based on a standard Gibbs sampling format in which we iteratively simulate elements of ψ and $\mathbf{X}_I$ from their conditional posteriors with all conditioning parameters fixed at their latest sampled values. In stationary models with no unit roots, the resulting reversibility of the model may be exploited to deduce the relevant conditional posterior for latent initial values $\mathbf{X}_I$. That is, for $t = 0, -1, \ldots, -(p-1)$ in turn, the initial values are sampled from the reverse time model $\phi(B^{-1})x_t = \tilde{\epsilon}_t$ or $x_t = \sum_{j=1}^{p} x_{t+j} + \tilde{\epsilon}_t$ where the $\tilde{\epsilon}_t$ are independent $N(\cdot|0, \sigma^2)$ variates. In cases of one or more unit roots, it turns out that the reverse time model can still be exploited to sample latent initial values, exactly as in stationary cases. This new and practically important result, which arises due to symmetries among the coefficients of component AR(1) and AR(2) processes with unit roots, is detailed in the Appendix and used in our analyses.

Missing observations are directly handled with extensions of the analysis as detailed in McCulloch and Tsay (1994). Write $\mathbf{X}_M$ for any subset of observations whose values are missing, assumedly at random, and redefine $\mathbf{X}$ to stand for the observed data only. The simulation analysis is easily extended to incorporate $\mathbf{X}_M$ as a set of additional latent variables, so that we are now simulating from the joint posterior $p(\psi, \mathbf{X}_I, \mathbf{X}_M|\mathbf{X})$. This simply involves an additional set of steps to sample conditional posterior distributions for each the elements of $\mathbf{X}_M$ in turn, and follows prior work by McCulloch and Tsay (1994). This is also relevant in the astronomical application of Section 4.

3 Component structure and spectral analysis

The basic decomposition result described in Section 2.1 provides a useful way of investigating and illuminating what are usually referred to as spectral characteristics. Each of the z_{tj} components is a stochastically varying quasi-periodic process with a characteristic frequency ω_j and damping factor r_j per unit of time. Inspection of graphs of the estimated processes z_{tj} provide insight into the time series structure. Very often, an appropriate model will decompose into several a_{tj} and z_{tj} processes of which only very few of the z_{tj} dominate in terms of amplitudes. It is frequently the case that these dominant components will

have higher moduli than the rest, consistent with the interpretation of the less dominant components as representing sources of noise.

Though spectral analysis using autoregressive models has a long history, the interpretation, in both spectral and time domains, of the component structure is a neglected topic and practical tool. For $|r_j| < 1$ the spectral density function of the latent z_{tj} process is

$$f_j(\omega) \propto |b_j + d_j e^{-i\omega}|^2 / |(1 - r_j e^{i(\omega_j - \omega)})(1 - r_j e^{-i(\omega_j + \omega)})|^2$$

as a function of frequency ω. This has a peak at ω_j that becomes more pronounced as $r_j \to 1$, degenerating into a "spike" at frequency ω_j when $r_j = 1$. The overall spectral density of the x_t process is

$$f(\omega) \propto \prod_{j=1}^{C} 1/|(1 - r_j e^{i(\omega_j - \omega)})(1 - r_j e^{-i(\omega_j + \omega)})|^2 \times \prod_{j=2C+1}^{p} 1/|1 - r_j e^{-i\omega}|^2$$

which may have several peaks depending on the relative moduli and amplitudes of the z_{tj} components. Traditional methods of autoregressive model fitting have led to concerns about the resulting estimates of spectra (Thomson 1990). Some of these questions related to parametric versus nonparametric spectral estimation are a side issue here; a parametric model that is demonstrably adequate in terms of its fit to the observed data is acceptable as a foundation for parametric spectral density estimation. A more important question is that associated with the difficulties of identifying sharp peaks or poles in spectra using parametric models. This is mainly due to the fact that standard methods of model parameter estimation constrain the roots to be less than one in modulus. As a result, such methods find it difficult to generate very high and sharp spectral peaks as such peaks correspond to regions of the parameter space that the estimation method is overtly, and often heavily, biased against. By design, our analysis allows unit roots generally, and hence unit moduli for the quasi-periodic latent component processes, through the prior specification. This leads to posteriors that will have meaningful mass at and near unit root values as the data demands, and resulting posterior distributions for the parametric spectral density functions will correspondingly have poles. In addition, and in line with our own preferences in summary inferences, the corresponding posterior distributions for the moduli and frequency parameters explicitly speak to the questions of whether or not a component is close to or on the stationarity boundary. This is illustrated in the following Section.

4 Analysis of variable star *S. CARINÆ* data

We analyse the time series of visual observations of light variations of the *S. Carinæ* variable star in the southern hemisphere sky. These data, from Baldwin and Thomson (1978), are of interest in exploring questions about periodicities in the light variations related to pulsations

of the variable star. The data represent 1189 observations constructed as consecutive 10-day averages of light intensities over a period of years. There are 40 points of missing data, indicated in the time series plot in Figure 1. From visual inspection of the time plot, the main period seems to be somewhere around 150 days (15 ten day means) which is equivalent to a frequency of 0.42 on the 10 day scale.

Our reported analysis assumes an autoregressive model of maximum order $p_+ = 60$ based on maximum numbers of complex and real components $C_+ = R_+ = 20$. We adopt hyperparameters $\beta = 1$ and $L = n/2$, and analyse the data after twice taking the natural logarithm of each observation and the subtracting the overall mean. The double log scale produces a transformed series in which the dominant periodic fluctuations appear more symmetric under reflection about the mean, so being more consistent with a linear time series model than the original data or other transforms we have explored. For this, purely empirical reason, we discuss analysis on this scale, though have explored analyses of the logged data too to confirm that there is little impact on the primary conclusions here about quasi-periodic component structure. Our analysis is based on a full posterior sample of size 5000, following exploration of burn-in and MCMC convergence questions. The analysis includes the 40 missing data values as latent variables and computes full joint posterior samples for these together with all model parameters and hyperparameters. Convergence checks included visual exploration of graphs of posterior sample series, monitoring of acceptance rates for Metropolis steps, and so forth.

The remaining figures summarise the posterior samples. The posterior distribution for model order p appears in Figure 2 together with the prior. The posterior gives positive mass to models of order 19-30 favouring models of order in the range 20-25, with the usual high degree of uncertainty about orders within this range, but evidently heavily discounts models of order less than 19 and higher than 30. Figure 3 shows the posterior distribution for the root structure of the model. For the complex roots, the model clearly favours 7 or 8, giving almost no probability to other values. The posterior for the number of real roots is more dispersed, as is typical since the real components are very low in amplitude and moduli and represent very high frequency noise. Some of the complex roots evidently relate to the meaningful component periodicities present in the data, as we discuss below.

Figure 4 displays the log-log data divided into 3 sections of about 400 observations each and with the posterior means for the 40 missing values substituted as actual observed values. These are indicated by the vertical dashed lines. The analysis evidently produces appropriate estimated values for these unobserved 10 day means.

Exploring the posterior requires, as discussed in Section 2.2, that we impose identifying constraints on the root structure of the autoregression. Here we order complex roots by frequencies, or wavelengths, as usual. We examine the model implications first by graphing

some of the estimated latent component processes derived from the analysis. Each of the complex roots corresponds to a quasi-periodic z_{tj} process that can be computed based on any specific set of parameters, imputed missing data and initial values. As a result we can compute the posterior means for each z_{tj} pointwise over time t. This is done and the results are graphed over time in Figure 5 for the the components related to the first several complex roots, ordered for identification in terms of frequencies. Each component is plotted on the same vertical scale as the data so that their amplitudes are directly comparable. Summaries of posteriors for the moduli and frequencies of these components appear in Figures 6 and 7 respectively, and the three graphs should be considered together in the following discussion. First, the low frequency component $j = 1$ has a modulus of about 0.94 and a frequency of near 0.11 which corresponds to a period of about 568.4 days; the posterior probability of a unit modulus is zero for this component. The variation over time of this sub-series is irregular and not evidently periodic, though well-explained by the low-frequency quasi-periodic model component. To our knowledge there is no scientific theory to explain this apparent signal in the data, though it does evidently reflect long term variability underlying the light series and has been noted in previous studies of this data (Baldwin and Thomson 1978).

The dominant, sustained periodic fluctuations in the data are evidently almost fully described by a single component, $j = 2$, of frequency very close to 0.420, or a period of about 149.8 days. This component has an associated posterior probability of 0.87 on a unit modulus, and so corresponds with high probability to a persistent harmonic component with little or no damping (West 1996; West and Harrison 1997, section 8.6.3). The next three complex roots, in order of increasing frequencies, have frequencies around 0.84, 1.26 and 1.67, respectively; these are very close to 2,3 and 4 times the frequency of the second component and so the corresponding latent components very likely represent harmonics of the fundamental and persistent component of frequency near 0.42. The estimated components have relatively low amplitudes over time, but do appear to contribute to the overall periodic patterns. One further component again of very low amplitude is apparently representing a mixture of the 5th and 6th harmonic frequencies; inferences at this high frequency/negligible amplitude level are clouded by noise and so the posteriors for the modulus and frequency are quite diffuse, though the latter does nicely span the range of frequencies corresponding to these two higher order harmonics. With the exception of the highly dominant fundamental harmonic, the additional components all have moduli that are clearly less than unity, and decay with in order of increasing frequency.

Figure 8 presents 100 samples from the posterior distribution for the theoretical autoregressive spectrum of the model, simply via evaluation of the spectrum at a posterior sample of 100 sets of the model parameters. The scale of the y-axis scale is decibels. As

discussed in the previous section, posterior point masses at unit moduli for complex components induce poles at the corresponding characteristic frequencies. As mentioned above, the dominant periodic component in this model analysis has high posterior probability of a unit root, and this is indicated in the figure through vertical lines at the sampled values of that frequency. For comparison, Figure 9 overlays the posterior mean of the spectrum on the traditional periodogram of the series with missing values set at their posterior means. This graph must be coupled with the summary posterior probability of 0.87 on a pole at the key frequency of around 0.42. We note that Carter and Kohn (1997) present a semiparametric approach to spectral estimation that combines harmonic regression with smooth spectral densities. These authors criticise the usage of traditional time series models, expressing concern about ad-hoc methods of model order selection and the introduction of potentially spurious features in the estimated spectrum when using high order models. We agree with these criticisms. However, it is the ways which the models are usually applied rather than the models themselves that are to be criticised. We have here demonstrated that these two issues are overcome within the standard autoregressive framework by the introduction of structured prior distributions that permit unit roots and suppress over-fitting through the inherent smoothness-prior structure and the resulting, appropriate model averaging over model order. We note that, as we can explore models of order up to and including the observed sample size in this framework, we are working in what is essentially the most general linear modelling framework.

We make some final comments on the Carinæ star data. In terms of inference on the basic periodicity, our analysis naturally broadly agrees with previous studies of Baldwin and Thomson (1978), Quinn and Thomson (1991), and Carter and Kohn (1997). There is an apparent and dominant fundamental frequency of near 0.42 and at least two harmonics. Our analysis goes further than previous work in identifying higher order harmonics and, particularly, in identifying the form and structure of the corresponding latent processes, as in Figure 5. Furthermore, we immediately identify the very low frequency component of period near 568 days that has eluded some researchers though was identified by Baldwin and Thomson (1978) through subsidiary spectral methods applied to residuals from initial models.

Final graphs in Figure 10 summarise posterior means of the innovations in the model. These suggest sightly heavier tails than normality, and a minor but suspicious degree of residual autocorrelation at or near the period of the dominant component cyclical component. As mentioned above, this high order model provides for approximation to essentially any linear process, so we conclude that this low level of residual structure, in both autocorrelations and in slight fat-tailedness, reflects non-linearity. Our inference is that, though the model and summary conclusions are basically sound, some minor perturbations may

be in order to account for this non-linearity. Referring to Baldwin and Thomson (1978) we note that there is some scientific rationale and interest in the possibility of long-term alterations in the fundamental frequencies of variable stars and that the detection of such would have implications for astrophysical theories of such stars. This suggests model modifications to allow small changes in autoregressive parameters over time, as in the class of time-varying autoregressions (Prado and West 1997; West and Harrison 1997 section 9.6), and these could easily account for both the slight heavy-tailedness and the residual autocorrelation structure. An alternative, or perhaps additional, extension would be to explore the use of heavy-tailed innovations errors, such as a Student-t error structure, which is easily accessible in the MCMC framework. Further development of the current prior modelling framework in these directions is a current research goal.

Acknowledgements

We are grateful to Robert Kohn for the provision of the version of the Carinæ star data set analysed here and to an anonymous referee for useful comments on an earlier version of the paper. This research was partially supported by the National Science Foundation under grants DMS-9704432 and DMS-9707914.

References

Baldwin, A.J. and Thompson, P.J. (1978) Periodogram analysis of S. Carinæ. *R. Astron. Soc. N.Z. (Variable Star Sect.)*, **6**, 31-5.

Carter, C., and Kohn, R. (1997) Semiparametric Bayesian inference for time series with mixed spectra. *J. Roy. Statist. Soc.*, (Ser. B), **59**, 255-268.

Huerta, G., and West, M. (1997) Priors and component structures in autoregressive time series models. To appear in *J. Roy. Statist. Soc., (Ser. B)*.

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), New York: Springer-Verlag.

Quinn, B.G., and Thomson, P.J. (1991) Estimating the frequency of a periodic function. *Biometrika*, **78**, 65-74.

Thomson, D.J. (1990) Time series analysis of Holocene climate data. *Phil. Trans. R. Soc. Lond. A*, **330**, 601-611.

West, M. (1996) Some statistical issues in Palaeoclimatology (with discussion). In *Bayesian Statistics 5*, (eds: J.O. Berger, J.M. Bernardo, A.P. Dawid and A.F.M. Smith), Oxford University Press.

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

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

Appendix: Model Reversibility in cases of Unit Roots

This appendix details the validity of the reverse time AR model for sampling latent initial values even in cases of one or more unit roots.

In general, let the AR(p) model have some number $a > 0$ unit roots and set $d = p - a$. Then we can write $\phi(B) = \mu(B)\psi(B)$ where $\mu(B)$ is the polynomial operator corresponding to the unit roots and $\psi(B)$ is the remaining polynomial operator corresponding to a strictly stationary AR(d) process. Furthermore, the filtered series $y_t = \mu(B)x_t$ follows this stationary AR(d) model and is therefore time reversible, i.e., $y_t = \psi(B^{-1})^{-1}\epsilon_t^*$ for all t for some innovations sequence $\epsilon_t^* \sim N(\cdot|0, \sigma^2)$. Then, given the observed data $x_1, \ldots, x_n$, we can directly compute y_t for $t > a$, and the reverse-time model for y_t implies that the initial latent initial values $y_a, y_{a-1}, \ldots$, can be simulated sequentially, backwards in time. Given these values, we can compute the required p initial values of the x_t process simply by inverting the defining equation $y_t = \mu(B)x_t$ for each t; to do this we need to understand the structure of $\mu(B)$ in more detail.

Suppose, with no loss of generality, that the d unit roots are comprised of e unit real roots and f pairs of unit-modulus complex roots with arguments ω_j, $(j = 1, \ldots, f)$. So $a = e + 2f$ and we have

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

where $\beta_j = 2\cos(\omega_j)$ for each $j = 1, \ldots, f$. That is,

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

where the $a - 1$ coefficients $m_1, m_2, \ldots, m_{a-1}$ are functions of the smaller number of quantities β_j, and the m_j exhibit what turn out to be important symmetries. In particular, it can be shown by induction on a that, for $j > a/2$,

$$m_j = (-1)^a m_{a-j}. \tag{1}$$

Now, from the defining identity $y_t = \mu(B)x_t$ we have $y_{t+a} = \mu(B)x_{t+a}$ for all t, or

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

and therefore

$$x_t = (-1)^a \sum_{j=1}^{a-1} m_j x_{t+a-j} + (-1)^{a+1} x_{t+a} + (-1)^a y_{t+a}. \tag{2}$$

Since $y_{t+a} = \psi(B^{-1})^{-1}\epsilon^*_{t+a}$ the final term in equation (2) is just $\psi(B^{-1})^{-1}\tilde{\epsilon}_t$ where the $\tilde{\epsilon}_t \sim N(0, \sigma^2)$ are independent. Now the coefficient symmetries in equation (1) imply that the first term in equation (2) can be re-expressed as

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

and therefore

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

This expression for x_t as a process is precisely of the form

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

or

$$\phi(B^{-1})x_t = \tilde{\epsilon}_t. \tag{3}$$

Evidently, equation (3) is the original AR(p) model in reversed time. As a result, the recursively defined conditional distributions of the latent initial values are precisely those obtained from the reverse time calculation even though the x_t series is not strictly stationary. This result may be exploited to trivially impute/sample latent initial values even when there are unit roots.

As an aside, note that this result is not applicable in cases of roots with moduli strictly greater than unity, as the critical symmetries in equation (1) are not obtained in such cases.

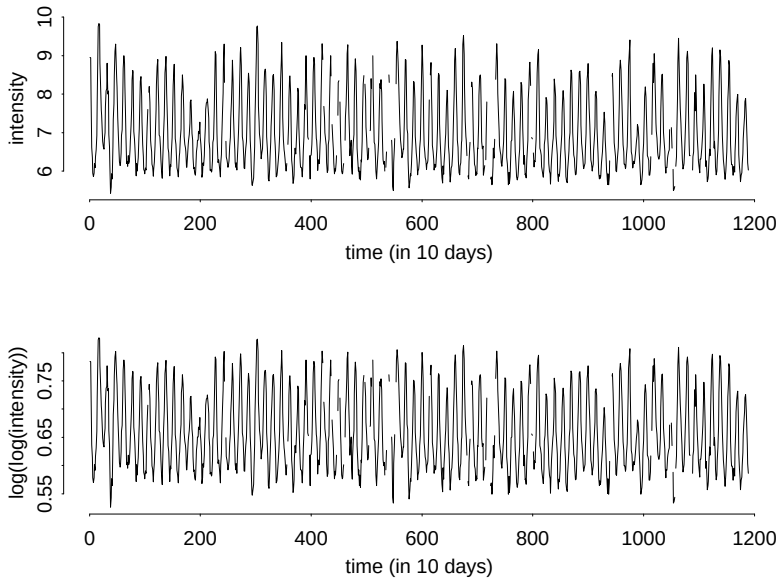

Figure 1: 1189 observations of light intensity from *S. Carinæ* star in both the original units and on a log-log scale.

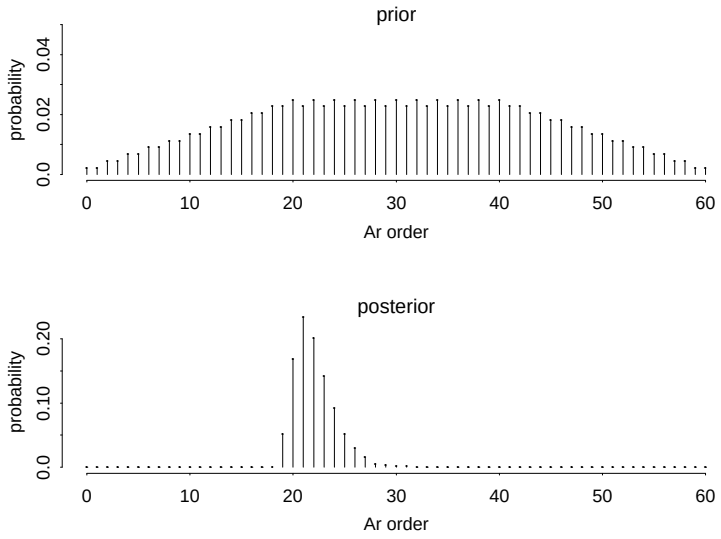

Figure 2: Prior and posterior distributions for model order p with $C_+ = 20$ and $R_+ = 20$.

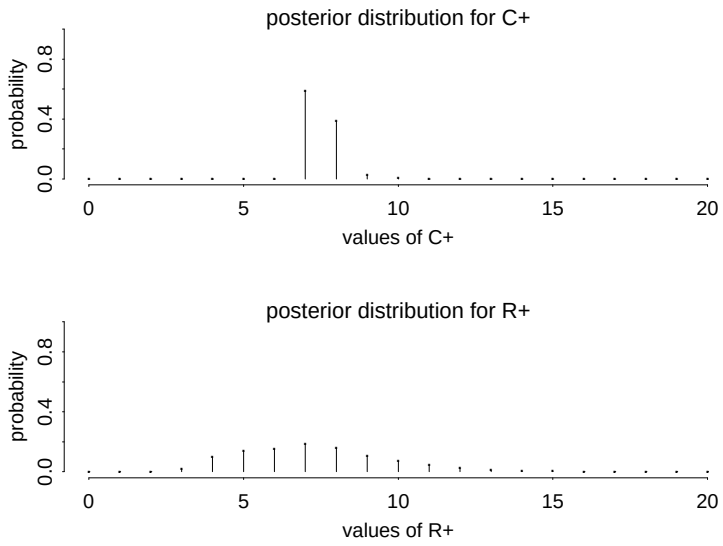

Figure 3: Posterior distributions for C and R.

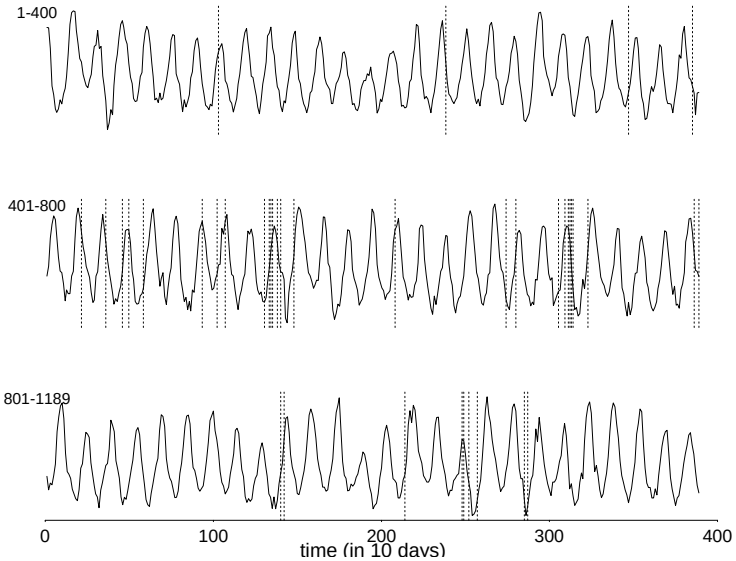

Figure 4: *S. Carinæ* star data on the log-log scale with missing values estimated by their posterior means, as marked by vertical dashed lines.

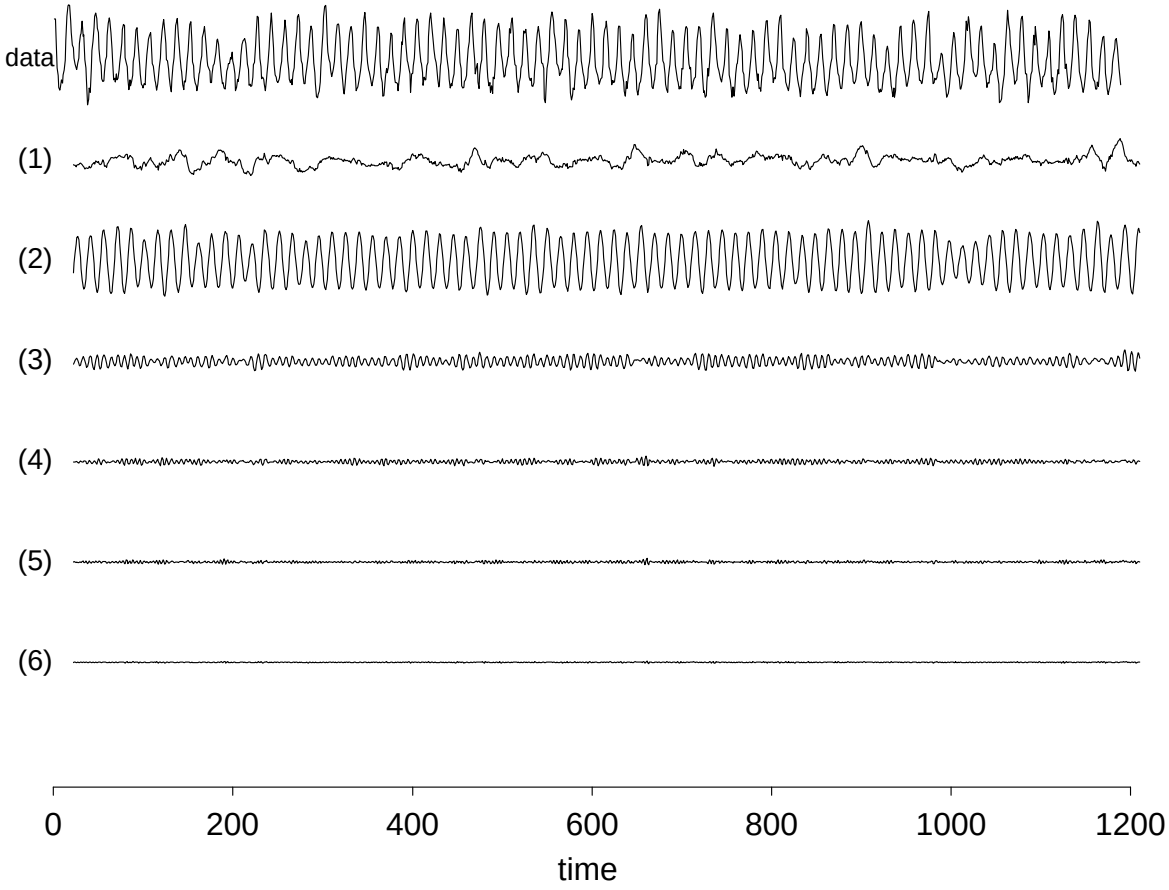

Figure 5: Star data and posterior means of the first six quasi-periodic latent components in order of increasing frequency, with component (1) that of lowest frequency.

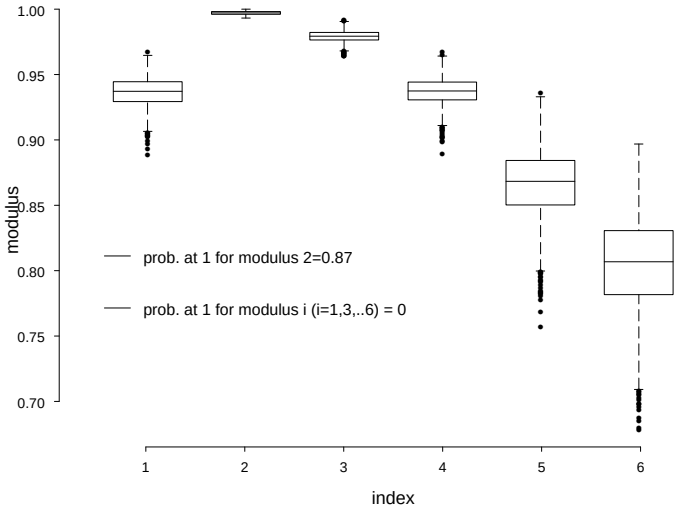

Figure 6: Boxplots of posterior samples for the moduli (excluding point mass at 1) of the six dominant complex roots ordered by frequencies.

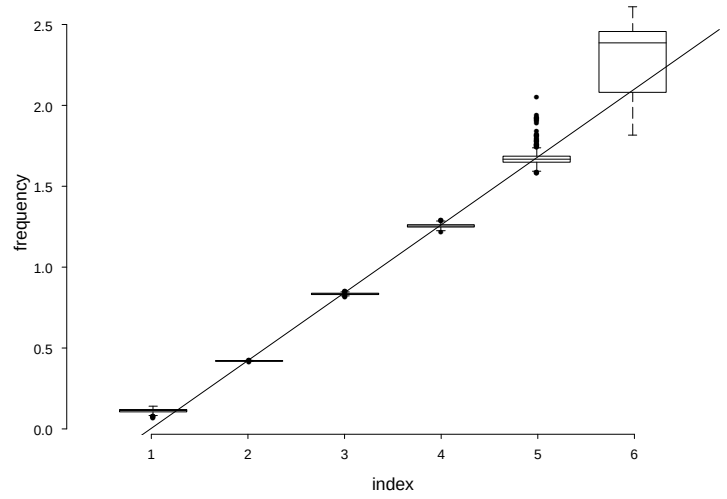

Figure 7: Boxplots of posterior samples for the frequencies of the six dominant complex roots ordered by frequencies. Exact harmonics of the key frequency at about 0.42 would lie exactly on the superimposed line.