Seasonality, Cycles and Unit Roots

Sune Karlsson*
Dept. of Economic Statistics

Mickael Salabasis†
Dept. of Economic Statistics
Stockholm School of Economics

Abstract

Inference on ordinary unit roots, seasonal unit roots, seasonality and business cycles are fundamental issues in time series econometrics. This paper proposes a novel approach to inference on these features by focusing directly on the roots of the autoregressive polynomial rather than taking the standard route via the autoregressive coefficients. Allowing for unknown lag lengths and adopting a Bayesian approach we obtain posterior probabilities for the presence of these features in the data as well as the usual posteriors for the parameters of the model.

Keywords: Bayesian; autoregression.
JEL codes: C11; C22; C15.

Stockholm School of Economics, P.O. Box 6501, SE-113 83 Stockholm, Sweden

*Department of Economic Statistics, Stockholm School of Economics.
Box 6501, 113 83 Stockholm, Sweden.

†Department of Economic Statistics, Stockholm School of Economics.
Box 6501, 113 83 Stockholm, Sweden.

1 Introduction

Seasons, cycles, and unit roots are dominant themes of time series econometrics. Having both important policy implications and a major impact on inference, these features are typically analyzed using the work horse autoregressive model. While inference on the autoregressive parameters is straightforward inference on the features of interest – seasons, cycles and unit roots – is more complicated since they depend on the roots of the autoregressive polynomial. To overcome this difficulty we utilize a novel model formulation and follow Huerta and West (1999*a*, 1999*b*) in writing the model in terms of and conduct inference directly on the roots. Adopting a Bayesian approach, the analysis is facilitated by assigning point mass priors on root configurations corresponding to the features of interest.

In addition to focusing directly on the roots of the process we adopt a flexible modelling framework. In particular we allow for uncertainty about the lag length, uncertainty about the number and nature of unit roots while allowing for polynomial trends in the data. The lag length is allowed to vary by letting individual roots be either null or non-null and thus contribute to the effective lag length. Allowing the number of positive unit roots to vary may cause trend parameters to become unidentified. We circumvent this problem by introducing an element of model selection, treating different combinations of number of positive unit roots and degree of the trend polynomial as distinct models.

This flexibility and ease of interpretation of the results comes at a cost. The main complication arises because the model is nonlinear in the roots of the characteristic polynomial and the posterior distribution is intractable. We use MCMC methods to analyze the posterior distribution with the reversible jump Markov Chain Monte Carlo algorithm of Green (1995) as a unifying computational framework.

The output of the Markov Chain contains a rich set of information on the posterior distribution. Posterior probabilities for the presence of particular features as well as the posterior distribution of the roots or the ordinary autoregressive parameters are easily obtained. In doing this we can either condition on specific features being present in the model (e.g. an AR(3) with one unit root and a constant term) or average over the different combination of features. In the latter case we are in effect performing Bayesian model averaging and the results are robust to model misspecification within the class of models considered here.

The present paper extends the work of Huerta and West in three directions. Firstly we add seasonal components which makes it possible to analyze the presence of seasonal unit roots and stationary seasonal features. Secondly, and perhaps most importantly, we generalize the model to allow for a polynomial trend of arbitrary degree. This is in contrast to Huerta and West who demean the data and exclude the constant term from the model. Their modelling framework is thus in effect only applicable to non-trending data. Thirdly, we give a different interpretation of the prior on the features of interest which makes it easier to specify the prior and avoids some potential pitfalls.

In related work Franses, Hoek and Paap (1997) propose a Bayesian analysis of seasonal unit roots based on the model formulation of Hylleberg, Engle, Granger and Yoo (1990). While computationally more demanding our approach is much more general, providing a richer set of information about the appropriate model and offering robustness against misspecification through model averaging. The approach of Dejong and Whiteman ((1991*b*), (1991*a*)) to unit root inference is similar in spirit to our work in that they consider the posterior distribution of the largest root of the characteristic polynomial. An important

difference is that DeJong and Whiteman do not model the roots directly, instead they model the AR parameters and solve for the roots given these parameters.

The organization of this paper is as follows. The modelling framework is introduced in Section 2 and a suitable prior is given in Section 3. The necessary tools for posterior analysis by MCMC are developed in Section 4. Section 5 applies the model to data on the Swedish GDP and Section 6 offers some closing remarks.

2 Modelling the features of interest

In most cases the features of interest correspond to single points in the parameter space. We will thus classify the roots accordingly and later on introduce a prior adapted to this classification. At this stage it suffices to note that a Bayesian treatment of features that correspond to a single point in the parameter space requires a prior that assigns point mass probabilities to each of them. Excluding explosive behavior, the roots of the autoregressive polynomial or, equivalently, the characteristic polynomial can be null, stationary or on the unit circle. By allowing for null roots the effective lag length can be varied and different lag lengths explored. For the active roots the classification into stationary roots and unit roots can be further refined by in each case considering roots at frequencies of special interest (typically seasonal) and roots at arbitrary frequencies. In each case the prior assigns a point mass to the particular feature thus permitting the calculation of posterior probabilities of the feature.

There is one feature of interest that does not fit directly into this framework and that is cycles of a general nature. An example is business cycles where the exact length of the cycle is unknown or may vary over time. Instead of considering a single frequency we need to focus on a range of frequencies. In this case it is not necessary to assign a point mass to the feature and it can be accommodated in the prior for roots at arbitrary frequencies. Still, modelling the roots directly makes inference on cycles of arbitrary length straightforward.

2.1 Stochastic and deterministic trends

Consider the autoregression of order p with a general polynomial trend of degree g

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=0}^{g} \delta_i t^i + \varepsilon_t, \tag{1}$$

where ε_t are iid $\mathcal{N}(0, \sigma_\varepsilon^2)$ innovations. Given (p, g) and an adequate number of initial values, this model may be analyzed within the standard framework of multiple regression.

The model formulation (1) is problematic in the sense that the interpretation of the model, in terms of the implied trend behavior, depends on the presence or absence of a real positive unit root in the lag polynomial; see Schotman and van Dijk (1991) and Lubrano (1995) for details. For each positive real unit root introduced in the lag polynomial, the interpretation of the trend parameters change. Alternatively, the model can be parameterized as

$$\phi(L)\left(y_t - \sum_{i=0}^{g} \beta_i t^i\right) = \varepsilon_t$$

where the order of the dynamics does not depend on the number of positive unit roots. Unfortunately, the underlying problem remains and is now manifest as local non-identification of trend parameters as a root approaches unity.

In line with our approach of modelling the features of interest directly we address the problem by explicitly treating each combination of the differencing order, d, and degree of the trend polynomial, g, as giving rise to distinct models with different interpretations. Decomposing $\phi(L)$ into $(1-L)^d \Psi(L)$ where the positive unit roots are excluded from $\Psi(L)$ we can rewrite the model as

$$\begin{aligned} (1-L)^d y_t &= \tau_t' \beta_{d,g} + \upsilon_t, \\ \Psi(L)\, \upsilon_t &= \varepsilon_t, \end{aligned} \tag{2}$$

where $\tau_t = (t^0, t^1, \ldots, t^g)'$ and g is implicitly redefined by subtracting d. This allow us to treat the stochastic and deterministic trends separately. In the posterior analysis we can then condition on a specific choice of d and g or average over them for a more robust analysis of the other features of interest.

While this solves the identification problem it is still the case that the data will be less and less informative about the constant term as the largest root of $\Psi(L)$ approaches unity. There is still the possibility of near non-identification and we will need to account for this in the prior specification.

2.2 Dynamics

Factoring the seasonal differencing polynomial

$$1 - L^S = (1-L)\left(1 + L + \ldots + L^{S-1}\right)$$

we obtain the seasonal unit roots as -1 and $\exp\left(\pm\frac{ik\pi}{S/2}\right)$, $k = 1, \ldots, S/2-1$ for S even. For the less common case of S odd the seasonal roots are instead given by $\exp\left(\pm\frac{ik\pi}{S/2}\right)$, $k = 1, \ldots, (S-1)/2$. In either case, we can add one real root and a set of complex roots to our collection of model features. Stationary seasonal dynamics is also of interest and we allow for complex roots at the seasonal frequencies with modulus less than unity. The set of roots is completed by adding stationary real roots and stationary complex roots with arbitrary modulus and frequency.

Since complex roots must appear in conjugate pairs the modelling exercise will be simplified if we treat real and complex roots separately. The lag polynomial $\Psi(L)$ will thus be taken to consist of a maximum of $\overline{r}$ real roots and $\overline{c}$ complex root pairs for a total maximum order of $\overline{p} = \overline{r} + 2\overline{c}$. Each of the $\overline{r}$ real roots may be either a null root, a stationary root or a negative unit root. Similarly each of the $\overline{c}$ complex pairs may be null, have an arbitrary frequency with arbitrary stationary modulus, seasonal frequency with stationary modulus or be a seasonal unit root.

It is convenient to parameterize the stationary lag polynomial $\Psi(L)$ in terms of the roots of the characteristic polynomials,

$$\Psi(L) = \prod_{j=1}^{\overline{r}} (1 - \rho_j L) \prod_{j=1}^{\overline{c}} \left(1 - 2m_j a_j L + m_j^2 L^2\right).$$

For the real roots the possible states are then null, $\rho_j = 0$, stationary, $0 < |\rho_j| < 1$ or a negative unit root, $\rho_j = -1$. For the complex pairs of roots, $m_j \exp(\pm i\omega)$, we have null roots,

$m_j = 0$, stationary roots, $0 < m_j < 1$ or unit roots, $m_j = 1$. For the non-null roots the modulus are combined with the seasonal frequencies, $\omega_j \in \{2k\pi/S,\ k = 1, \ldots, S/2 - 1\}$ or an arbitrary frequency, $0 < \omega_j < \pi$ to obtain the full set of complex roots. For ease of notation we let $a_j = \cos \omega_j$.

2.3 The likelihood

The parametrization in terms of the roots simplifies the identification, interpretation, and introduction of interesting features. However, while the general form of the likelihood does not change, the root parameters appear in a non-trivial multiplicative way. Conducting joint inference on the full set of parameters is thus complicated. On the other hand, the likelihood is amenable to conditional inference and we can divide the problem into smaller, simple, parts that – provided this structure is maintained by the prior – naturally leads to a MCMC algorithm for analyzing the posterior distribution. In this process, filtering plays an important role.

Given the differencing order, d, and the degree of the polynomial trend, g, define the trend vector $\tau_t = (t^0, t^1, \ldots, t^g)$ and $\widetilde{y}_t = (1 - L)^d y_t$. Then, for inference on the trend coefficient vector, conditional on the roots of $\Psi(L)$ the model in (2) reduces to

$$\begin{aligned} y_t^* &= \tau_t^{*\prime} \beta_{d,g} + \varepsilon_t, \\ y_t^* &= \Psi(L)\, \widetilde{y}_t,\ \tau_t^{*\prime} = \Psi(L)\, \tau_t' \end{aligned} \tag{3}$$

a standard linear regression model.

Conversely, for inference on the roots of the lag polynomial $\Psi(L)$, conditional on $\beta_{d,g}$, let $u_t = (\widetilde{y}_t - \tau_t' \beta_{d,g})$ be the centered appropriately differenced data. For a real root ρ_i, conditional on all else, define

$$z_t = \prod_{j \neq i} (1 - \rho_j L) \prod_{j=1}^{\overline{c}} \left(1 - 2m_j a_j L + m_j^2 L^2\right) u_t \tag{4}$$

and conditional inference on ρ_i can be conducted in

$$(1 - \rho_i L)\, z_t = \varepsilon_t,$$

an autoregressive process of order one with known variance σ_ε^2. Similarly for a complex pair, $m_i \exp(\pm i\omega_i)$, define

$$z_t = \prod_{j=1}^{\overline{r}} (1 - \rho_j L) \prod_{j \neq i} \left(1 - 2m_j a_j L + m_j^2 L^2\right) u_t. \tag{5}$$

Then, conditional on all but the complex pair the model reduces to

$$\left(1 - \phi_1 L - \phi_2 L^2\right) z_t = \varepsilon_t,\ \phi_1 = 2m_i a_i = 2m_i \cos \omega_i,\ \phi_2 = -m_i^2,$$

an autoregressive process of order two with parameter restrictions.

A remaining issue is the treatment of the initial values. We can either take them as given and use the likelihood for $y_{\overline{p}+d+1}, \ldots, y_T$ conditional on $y_1, \ldots, y_{\overline{p}+d}$ or use the full likelihood for $y_1, \ldots, y_T$. The former has the advantage of simplicity but can lead to a substantial loss of information if $\overline{p} = 2\overline{c} + \overline{r}$ is large relative to T. Instead, we prefer the latter approach and marginalize out the unobserved initial values $y_{1-\overline{p}-d}, \ldots, y_0$, by treating them as latent variables to be simulated together with the parameters in the Markov chain.

3 The Prior Structure

The overall prior structure follows from a few reasonable independence assumption. We take the seasonal and cyclical behavior to be a priori independent of the trend properties. In addition, the innovation variance, σ_ε^2, plays a different role in autoregressive models than in standard regression models. In particular it does not influence how informative the data is about the parameters in the autoregressive polynomial and the usual conditioning on σ_ε^2 does not simplify the prior specification. Decomposing the prior on the trend parameters β, d and g as $\pi\left(\beta_{d,g}|\sigma_\varepsilon^2, d, g\right)\pi\left(d, g\right)$ we obtain the prior as

$$\pi\left(\theta, \beta_{d,g}, d, g, \sigma_\varepsilon^2\right) = \pi\left(\theta\right)\pi\left(\beta_{d,g}|\,\sigma_\varepsilon^2, d, g\right)\pi\left(d, g\right)\pi\left(\sigma_\varepsilon^2\right) \tag{6}$$

where the characteristic roots of $\Psi\left(L\right)$ are collected in

$$\theta = \left(\rho_1, \ldots, \rho_{\overline{r}}, m_1, a_1, \ldots, m_{\overline{c}}, a_{\overline{c}}\right).$$

3.1 Stochastic and deterministic trends

In most cases only a few possible values for d and g are considered, i.e. at most two unit roots or at most a quadratic trend. This leads to a limited number of combinations $(d, g) \in \left(0, \ldots, \overline{d}\right) \times \left(-1, \ldots, \overline{g}\right)$ where $g = -1$ corresponds to a model with no constant term or trends. In addition, some of these might be ruled out a priori as unreasonable. Assigning prior probabilities is thus fairly straightforward and if nothing else a uniform prior on the possible combinations can be used.

3.2 Lag polynomial prior

Turning to the characteristic roots of $\Psi\left(L\right)$ we note that the polynomial is invariant to permutations of the roots. Consequently we treat the roots symmetrically and specify identical and independent priors for the the real and complex roots respectively. We write the prior on θ as

$$\pi\left(\theta\right) = \prod_{j=1}^{\overline{r}}\pi\left(\rho_j\right)\prod_{k=1}^{\overline{c}}\pi\left(m_k, a_k\right) \tag{7}$$

where $\pi\left(\rho_j\right)$ is the prior on an arbitrary real root and $\pi\left(m_k, a_k\right)$ the prior on an arbitrary complex root. This prior structure is close to the one used by Huerta and West (1999*a*, 1999*b*) with obvious modifications to cater for the differences in model set up.

3.2.1 Real roots

For real roots we distinguish between three states, null roots, stationary roots and negative unit roots. Corresponding to this the prior is a mixture over the three states,

$$\pi\left(\rho\right) \sim w_0^r I_{\rho=0} + w_s^r f_{|\rho|<1}\left(\rho\right) + w_{-1}^r I_{\rho=-1}, \tag{8}$$

where the weights $\{w_i^r\}$ sum to one, $I_{\rho=k}$ is the usual indicator function, and $f_{|\rho|<1}\left(\rho_j\right)$ is a continuous distribution with support on the stationarity region.

When interpreting the prior the primary distinction is between null roots and non-null roots which contribute to the effective lag length. The non-null roots can in turn be either

stationary or a unit root. This hierarchical structure is useful when thinking about the weights. We suggest choosing the weights as

$$w_0^r = p_0,\ w_s^r = (1-p_0)\,p_s,\ w_{-1}^r = (1-p_0)\,(1-p_s)$$

where p_0 is the a priori expected proportion of null roots and p_s is the a priori expected proportion of stationary roots given that the root is active. A natural uninformative prior specification given this structure is $p_0 = 1/2$, $p_s = 1/2$. Note that this differs from the uninformative prior, $w_1^r = w_2^r = w_3^r = 1/3$, obtained by considering (8) directly. The latter is essentially the prior used by Huerta and West (1999*b*). Given $\overline{r}$, the number of real roots, the choice of weights induces a multinomial prior on the number of roots of the different types. In particular, the induced prior on the contribution of the real roots to the overall lag length is binomial, $Bin\,(\overline{r}, 1-w_1^r)$, and it is important to keep this in mind when choosing $\overline{r}$.

To complete the specification of the prior we need to specify $f_{|\rho|<1}\,(\rho)$, the prior on stationary roots. Convenient choices are a uniform prior on $(-1,1)$ and a truncated normal distribution. Both these priors are problematic in the sense that the posterior for the constant term in the deterministic trend will be dominated by the prior as $\rho \to 1$. In the context of a model with only a constant term Schotman and van Dijk (1991) propose dealing with this problem by making the prior for the constant more concentrated around the initial condition, see Schotman (1994) for a detailed discussion of these issues. The prior of Schotman and van Dijk has the disadvantage that it is data dependent and it is unclear how this should be generalized to trend models of higher order. Instead we take the simple route of specifying a uniform prior on the restricted range $(-1, 1-\delta)$ for a small value of δ. In connection with this it is important to note that this does note rule unit roots as these are modeled together with the deterministic trend. Instead we view this as driving a wedge between trend stationary and unit root dynamics and forcing the model to clearly distinguish between them.

3.2.2 Complex roots

The prior for the complex conjugate pairs can be constructed in a similar manner. A root is null with probability p_0 or otherwise non-null. If it is non-null, it may be stationary with probability p_s or on the unit circle. In addition, non-null roots may have a frequency of special interest with probability $p_{\omega_i^*}$ or have an arbitrary frequency with probability $1 - \sum p_{\omega_i^*}$. In the case of a single frequency of special interest, ω^*, we would have

$$\begin{aligned}\pi\,(m, a = \cos\omega) \sim w_0^c I_{m=0} + w_{s,\omega^*}^c f_{m,\omega=\omega^*} \\ + w_{s,\omega}^c f_{m,\omega} + w_{1,\omega^*}^c I_{m=1,\omega=\omega^*} + w_{1,\omega}^c f_{m=1,\omega}\end{aligned} \quad (9)$$

with

$$\begin{aligned} w_0^c = p_0,\ w_{s,\omega^*}^c = (1-p_0)\,p_s p_{\omega^*},\ w_{s,\omega}^c = (1-p_0)\,p_s\,(1-p_{\omega^*}) \\ w_{1,\omega^*}^c = (1-p_0)\,(1-p_s)\,p_{\omega^*},\ w_{1,\omega}^c = (1-p_0)\,(1-p_s)\,(1-p_{\omega^*}) \end{aligned}$$

where p_0 is the a priori expected proportion of null roots, p_s the proportion of stationary roots given that the root is non-null and p_{ω^*} the proportion of non-null roots with the special frequency ω^*.

Apart from the structure imposed on the prior weights, this is a straightforward extension of the prior used by Huerta and West (1999*b*). It should be clear that the structure

plays a larger role with the complex roots than with the real roots. It greatly reduces the number of prior parameters to specify and implies a more reasonable default, uninformative, prior with $p_0 = 1/2$, $p_s = 1/2$ and $p_{\omega^*} = \frac{1}{n^*+1}$ where n^* is the number of special frequencies. Recall that the implied prior for the number of non-null complex roots is $Bin\left(\overline{c}, 1 - w_0^c\right)$. It is clear that the choice $w_0^c = \frac{1}{2(n^*+1)+1}$ inspired by considering (9) directly can lead to an implausible large number of complex roots if, as prudence might dictate, a large value of $\overline{c}$ is used.

To complete the specification we need to specify the continuous components. Following Huerta and West (1999*b*) a joint uniform prior on the implied autoregressive parameters in (5) is used. For the component corresponding to a stationary cycle at an arbitrary frequency, $f_{m,\omega}$, this results in

$$m \sim Beta\left(3,1\right), a \equiv \cos\omega \sim U\left(-1, \bar{a}\right).$$

where $\bar{a} = \cos\underline{\omega}$. The upper bound depends on the minimum observable frequency $\underline{\omega}$ which in turn depends on the length of the time series in a trivial way. This leads to a truncated bivariate normal posterior and potential simplifications in the posterior simulation for the implied parameters in the time domain. For the other continuous components added by the presence of special frequencies, a similar argument leads to the prior $f_{m,\omega=\omega^*} \sim Beta\left(2,1\right)$ for the modulus with ω fixed at a special frequency. Finally, for unit roots with arbitrary frequency, the same uniform prior $a \equiv \cos\omega \sim U\left(-1, \bar{a}\right)$ is used for the frequency.

3.3 Polynomial trend prior

By isolating the positive unit roots and rewriting the model as (2) a singularity in the conditional likelihood (3) for $\beta_{d,g}$ is avoided at the cost of making the interpretation of $\beta_{d,g}$ depend on the differencing order, d. Specifying a general prior for $\beta_{d,g}$ is thus difficult unless we make it uninformative or vague. A uniform, improper, prior is in fact possible after the removal of the singularity but this choice is problematic if we view the choice of d and g as a model selection exercise. Instead we use a g-prior with the scaling factor selected to make the prior vague. That is,

$$\beta_{d,g} | \, d, g \sim \mathcal{N}_{g+1}\left(0, \sigma_\varepsilon^2 M^{-1}\right)$$

where $M = gX'X$ and X the trend matrix with rows $\tau_t' = \left(t^0, t^1, \ldots, t^g\right)$.

3.4 The innovation variance

For the innovation variance, σ_ε^2, we use the standard inverse gamma prior,

$$\sigma_\varepsilon^2 \sim \mathcal{IG}\left(a_0, b_0\right),$$

with a_0 and b_0 selected to make the prior proper but vague. This leads to an inverse gamma full conditional posteriors and simplifies the MCMC computations.

4 The Posterior Simulation

4.1 Implementation

Exploration of the model space requires occasional changes in the dimension or content of the parameter vector. The dimension changing moves are handled using the reversible

jump MCMC algorithm of Green (1995). This entails the construction of moves between the possible states together with their balancing counterparts. Some additional notation is needed when discussing the moves. Let the subscripts c and p denote current and proposed quantities. Collecting all current parameters in $\xi_c = [\theta_c, \beta_c, d_c, g_c, \sigma_c^2]$, the prior of the current state is summarized by $\pi_c = \pi(\xi_c)$. When appropriate, let j_{cp} be the probability of selecting a move that attempts a transition from the current state to the proposed state

The reversible jump algorithm solves the problem with changing dimensions by completing the parameter spaces. Let u_c and u_p be parameter vectors satisfying the dimension matching requirement $\dim(\xi_p) + \dim(u_c) = \dim(\xi_c) + \dim(u_p)$ and $(u_c, \xi_p) = g(\xi_c, u_p)$ a bijection. A proposal parameter vector, ξ_p, for the proposed state is then obtained by drawing u_p from a suitable distribution q_p and applying the transformation. As this is a Metropolis-Hastings step the corresponding, imaginary, operations are conducted for the balancing move from the proposed to the current state when forming the acceptance probability,

$$\alpha = \min\left\{1, \frac{L(\mathbf{y}|\xi_p)}{L(\mathbf{y}|\xi_c)} \frac{\pi_p}{\pi_c} \frac{j_{pc} q_c(u_c)}{j_{cp} q_p(u_p)} \left|\frac{\partial g(\xi_c, u_p)}{\partial(\xi, u)}\right|\right\},$$

where j_{cp} and j_{pc} are the probabilities of proposing a move that attempts a transition from the current to the proposed state and vice versa.

When $\dim(\xi_c) > \dim(\xi_p)$ it is common to take u_p to be empty and $\dim(u_c) = \dim(\xi_c) - \dim(\xi_p)$. The proposal would not involve any sampling although the balancing move is still conducted as if u_c was a random quantity. In contrast with this we will find it convenient to complete both parameter vectors in some cases. In practice, the proposals involve only a subset of the parameters and the acceptance probabilities simplify considerably by using the appropriate conditional likelihood and exploiting the structure of the prior.

For the particular problem of interest, the structure of the model suggests two major subsets, one containing the trend parameters (d, g) and another containing the collection of all roots. We use four types of moves to explore the (d, g) space. The first (T1) is a local move that holds (d, g) constant and only updates the coefficients of the trend polynomial. The second move (T2) changes the degree of the trend polynomial, $g_p = g_c \pm 1$, while holding the differencing order constant. The third move (T3) is the corresponding move for the differencing order. The final move (T4) change both quantities in such a way as to keep the total order of the trend constant, $d_p + g_p = d_c + g_c$. With the exception of T1, all moves require the execution of a reversible jump.

When exploring the dynamic properties, real and complex roots are treated separately. For real roots, as discussed in Section 3.2.1, a uniform prior over the stationary region for the continuous component leads to updates with a Gibbs step. For the complex roots the updates are more complicated and four different moves are used. Corresponding to the major states of the roots, M1 moves between null and stationary roots and M2 moves between stationary roots and roots on the unit circle. Complementing this, M3 explores the modulus and frequency of stationary roots and M4 explores the frequency of the complex unit roots. Both major moves are reversible jumps. While M3 in theory could be updated with a Gibbs step, the resulting complicated truncated bivariate normal prompts the use of a Metropolis-Hastings step. Contrasting, in M4 it is possible to update the frequency of a pair at the boundary using a procedure similar to that for real roots.

The Markov chain proceeds by cycling through the steps in Algorithm 1. In what follows let Y be the vector of original time series measurements, including the appropriate

Algorithm 1 Structure of the Markov chain

1. Update the trend properties. Select a new state (d_p, g_p) and attempt a transition to the state using the relevant move in Section 4.2
2. Update the lag polynomial (Section 4.3). Establish a cycle of length $l = \overline{r} + \overline{c}$. In each iteration sample a root index at random
   (a) If the selected index is associated with a real root, update it using the full conditional posterior in Section 4.3.1.
   (b) If the selected index is associated with a complex pair, query the current status and attempt to update it using one of four moves detailed in Section 4.3.2.
3. Update the variance σ_ε^2 using the full conditional posterior in Section 4.4.
4. Sample new initial values using the procedure detailed in Section 4.5.

number of latent initial values, and X the matrix of appropriately defined trend variables.

4.2 Updating the trend properties

Propose a new state (d_p, g_p) with equal probability from $\{(d,g) : (d_c, g_c),\ (d_c \pm 1, g_c),\ (d_c, g_c \pm 1),\ (d_c \pm 1, g_c - d + d_c),\ 0 \leq d \leq \overline{d},\ -1 \leq g \leq \overline{g}\}$. Due to the endpoint restrictions the available states vary and the proposal probabilities are $j = 1/3$ for $(d_c, g_c) = (0, -1)$ or $\left(\overline{d}, \overline{g}\right)$, $j = 1/4$ for $(d_c, g_c) = (0, \overline{g})$ or $\left(\overline{d}, 0\right)$, $j = 1/5$ for $d_c \in \left\{0, \overline{d}\right\}$ and $-1 < g_c < \overline{g}$ or $g_c \in \{-1, \overline{g}\}$ and $0 < d_c < \overline{d}$ and $j = 1/7$ otherwise. Given the proposed state a transition is attempted using one of the moves T1 - T4. Throughout this section we condition on the current lag polynomial and innovation variance and for simplicity of notation we omit θ_c and σ_c^2 from the conditioning variables.

Although we consider four different moves they all fit in a common generic template. The key to the simplicity of the updates is that we draw the proposal for β directly from the full conditional posterior and treat the current value as a draw from the full conditional posterior. That is $q_p(\beta_p) = p(\beta|Y, d_p, g_p)$ and $q_c(\beta_c) = p(\beta|Y, d_c, g_c)$ and $g(\cdot,\cdot)$ is the identity function with unit Jacobian.

To illustrate, for the current model, let

$$E_c = \Psi^c(L)\,\widetilde{Y}_c, \quad \widetilde{Y}_c = (1-L)^{d_c} Y,$$

be the filtered appropriately differenced series, and filter the relevant section of the trend matrix X to obtain

$$D_c = \Psi^c(L)\,X_c.$$

With the conditional normal structure, the normal prior leads to the standard full conditional posterior

$$\beta_c \sim \mathcal{N}_{g_c+1}\left(\overline{\beta}_c, \Omega_c\right)$$

where

$$\overline{\beta}_c = \Omega_c\left[\sigma_\varepsilon^{-2} D_c' E_c\right],\ \Omega_c = \sigma_\varepsilon^2\left[D_c' D_c + M_c\right]^{-1}.$$

Collecting all factors in the acceptance probability that are associated with the current value, an expression for the current contribution simplifies to

$$\frac{j_{pc}q_c\left(\beta_c\right)}{L\left(\mathbf{y}|\xi_c\right)\pi_c} = j_{pc}\kappa_c\sqrt{\frac{1}{|\sigma_\varepsilon^2 M_c|\,|\Omega_c|}}\exp\left\{\frac{1}{2}\left[\sigma_\varepsilon^{-2}E_c'E_c - \overline{\beta}_c'\Omega_c^{-1}\overline{\beta}_c\right]\right\},$$

where κ_c is used to collect the scale factors from the normal likelihood, prior and proposal. From a similar exercise for the proposed state we get the analogous proposal quantities E_p, D_p, $\overline{\beta}_p$, and Ω_p. Some matrix algebra yields a simplified generic expression for the acceptance probability as

$$\alpha_T = \min\left\{1, \frac{j_{pc}}{j_{cp}}\sqrt{\frac{|\Omega_p|\,|\sigma_\varepsilon^2 M_p|}{|\Omega_c|\,|\sigma_\varepsilon^2 M_c|}}\frac{\exp\left\{-\frac{1}{2}\left[\sigma_\varepsilon^{-2}E_p'E_p - \overline{\beta}_p'\Omega_p^{-1}\overline{\beta}_p\right]\right\}}{\exp\left\{-\frac{1}{2}\left[\sigma_\varepsilon^{-2}E_c'E_c - \overline{\beta}_c'\Omega_c^{-1}\overline{\beta}_c\right]\right\}}\right\}. \tag{10}$$

Depending on the selected move, this expression may simplify further.

Move T1 (Updating the polynomial trend parameters)

When the differencing order and the trend degree are the same, the current and proposal quantities in (10) are all equal. The acceptance probability simplifies to $\alpha_{T1} = 1$ and this is equivalent to a Gibbs step. Note the special case $g_c = g_p = -1$ where no computations at all are necessary.

Move T2 (Updating the trend polynomial degree)

With $(d_p, g_p) = (d_c, g_c \pm 1)$ the current and proposal models share the same centered differenced series and the acceptance probability simplifies to

$$\alpha_{T2} = \min\left\{1, \frac{j_{pc}}{j_{cp}}\sqrt{\frac{|\Omega_p|\,|\sigma_\varepsilon^2 M_p|}{|\Omega_c|\,|\sigma_\varepsilon^2 M_c|}}\frac{\exp\left\{\frac{1}{2}\overline{\beta}_p'\Omega_p^{-1}\overline{\beta}_p\right\}}{\exp\left\{\frac{1}{2}\overline{\beta}_c'\Omega_c^{-1}\overline{\beta}_c\right\}}\right\}.$$

The expression simplifies further if either the current or proposed model implies a trend free model. For instance, assuming $g_c = -1$ we get

$$\alpha_{T2} = \min\left\{1, \frac{j_{pc}}{j_{cp}}\sqrt{|\Omega_p|\,|\sigma_\varepsilon^2 M_p|}\exp\left\{\frac{1}{2}\overline{\beta}_p'\Omega_p^{-1}\overline{\beta}_p\right\}\right\}.$$

Move T3 (Updating the differencing order)

With $(d_p, g_p) = (d_c \pm 1, g_c)$ the filtered trend matrix, D, and the prior precision are unaffected with $\Omega_c = \Omega_p$ and the acceptance probability simplifies to

$$\alpha_{T3} = \min\left\{1, \frac{j_{pc}}{j_{cp}}\frac{\exp\left\{-\frac{1}{2}\left[\sigma_\varepsilon^{-2}E_p'E_p - \overline{\beta}_p'\Omega_p^{-1}\overline{\beta}_p\right]\right\}}{\exp\left\{-\frac{1}{2}\left[\sigma_\varepsilon^{-2}E_c'E_c - \overline{\beta}_c'\Omega_c^{-1}\overline{\beta}_c\right]\right\}}\right\}.$$

Note how the major simplification is the cancelling of the ratio of scale factors. If there is currently no active trend model, that is $g_c = g_p = -1$, the expression simplifies to

$$\alpha_{T3} = \min\left\{1, \frac{j_{pc}}{j_{cp}}\exp\left\{-\frac{1}{2}\left[\sigma_\varepsilon^{-2}\left(E_p'E_p - E_c'E_c\right)\right]\right\}\right\}$$

Move T4 (Updating the trend degree and the differencing order)

Setting $(d_p, g_p) = (d_c \pm 1, g_c - d_p + d_c)$ no further simplifications are achieved and the generic expression in (10) is used so that $\alpha_{T4} = \alpha_T$.

Note that the acceptance probabilities do not depend on the current or proposed value of β. This suggests the following strategy to improve the mixing properties of the chain. Unless Move T1 is actually selected, sample the next trend state using the appropriate acceptance probability. Then, regardless of the result, sample a new coefficient vector from the relevant full conditional posterior necessary for updating the lag polynomial. Essentially, the updating sequence in Step 1 of Algorithm 1 then always ends with the execution of Move T1.

4.3 Updating the lag polynomial

Conditional on all else, and in particular the trend properties d, g and $\beta_{d,g}$, the data can be filtered as outlined in Section 2.3 to isolate the contribution of a single root. It is then straightforward to obtain the full conditional posterior for the root up to a scaling factor. To improve the mixing properties of the chain we update the roots in random order. Draw $l = \overline{r} + \overline{c}$ integers with replacement from $1, \ldots, \overline{r} + \overline{c}$ and update the roots in this order.

4.3.1 Sampling real roots

Assuming a real root ρ_i is selected for updating, define

$$U = \widetilde{Y} - X_c \beta_c, \quad \widetilde{Y} = (1 - L)^{d_c} Y,$$

the demeaned appropriately differenced series. Using (4), filter U to obtain the dependent variable Z and explanatory variable Z_{-1} for the AR(1). With a uniform prior for the continuous component in (8) , the full conditional posterior is

$$\rho_i | \, Y, \xi_c \diagdown \rho_i \propto \left(w_0^r I_{\rho=0} + w_s^r \frac{1}{2} + w_{-1}^r I_{\rho=-1} \right) \exp \left\{ -\frac{1}{2s^2} \left(\hat{r} - \rho_i \right)^2 \right\},$$

where

$$\hat{r} = \left(Z_{-1}' Z_{-1} \right)^{-1} Z_{-1}' Z \quad s^2 = \sigma_\varepsilon^2 \left(Z_{-1}' Z_{-1} \right)^{-1},$$

are the usual least squares quantities. Computing the scale factor, the full conditional posterior state probabilities for a null, stationary, and real negative unit root are given by

$$p_i = \frac{t_i}{\sum_{k=1}^{3} t_k},$$

where

$$\begin{aligned} \rho_i = 0: \quad & t_1 = w_0^r \exp \left\{ -\tfrac{1}{2s^2} \left(\hat{r} - 0 \right)^2 \right\}, \\ |\rho_i| < 1: \quad & t_2 = \tfrac{w_s^r}{2} \textstyle\int_\rho \exp \left\{ -\tfrac{1}{2s^2} \left(\hat{r} - \rho \right)^2 \right\} = \tfrac{w_2^r \cdot \kappa}{2} \left(2\pi s^2 \right)^{1/2} \\ \rho_i = -1: \quad & t_3 = w_{-1}^r \exp \left\{ -\tfrac{1}{2s^2} \left(\hat{r} - (-1) \right)^2 \right\}. \end{aligned}$$

In the expression for t_2,

$$\kappa = \Phi \left(\frac{1 - \hat{r}}{s} \right) - \Phi \left(\frac{-1 - \hat{r}}{s} \right)$$

is a scale factor due to the restriction $\rho_i \in (-1, 1)$. Sampling a new state, if the next state corresponds to a stationary root the actual value is sampled from

$$\rho \sim \mathcal{TN} \left(\rho | \hat{r} = \left(Z_{-1}' Z_{-1} \right)^{-1} Z_{-1}' Z \; s^2 = \sigma_\varepsilon^2 \left(Z_{-1}' Z_{-1} \right)^{-1}, -1, 1 \right),$$

a normal distribution truncated to the stationary region.

4.3.2 Sampling complex pairs

Assuming a complex pair $m_c \exp(\pm i\omega_c)$ is selected for updating, define

$$U = \widetilde{Y} - X_c\beta_c, \quad \widetilde{Y} = (1-L)^{d_c} Y,$$

the demeaned appropriately differenced series. Using (5), filter U to obtain the filtered sequence Z which is an AR(2) with non-linear parameter restrictions. While possible, a Gibbs step would be complicated and time consuming to execute. Instead, we use Metropolis-Hastings steps to move between the possible states for the complex roots.

Applying the reversible jump, move probabilities must be selected. As the suggested moves are tailored for transitions between particular root states, the jump probabilities will the very least depend on the current state. In particular, if the root selected for updating is at the origin, with probability j_{ns} attempt to introduce a stationary root using Move M1 and otherwise do not update the root. If the root selected for updating is instead stationary we may attempt Moves M1, M2, or M3, with probability j_{sn}, j_{su}, and $j_{ss} = 1 - j_{sn} - j_{su}$ respectively. In turn, this amounts to attempting to deactivate the root by placing it at the origin, moving it to the boundary and introducing a persistent cycle or seasonal component, or simply updating it. Finally, if the root selected is currently placed at the unit circle, an attempt to move it into the stationary region using Move M2 is performed with probability j_{us}. Otherwise, an attempt to update the frequency using Move M4 is made with probability $j_{uu} = 1 - j_{us}$.

Throughout we will, without loss of generality, assume that there is a single special frequency, $a^* = \cos\omega^*$.

Move M1 (Transitions between null and stationary roots)

Assume the complex pair is currently at the origin. To sample a proposal, from the filtered sequence obtain the dependent variable Z and the explanatory variables Z_{-1} and Z_{-2} for the AR(2). Calculate ordinary least squares quantities

$$W = [Z_{-1}, Z_{-2}], \quad \hat{\phi} = (W'W)^{-1} WZ, \quad \Omega = \sigma_\varepsilon^2 (W'W)^{-1}.$$

Begin by sampling a proposal for the modulus by sampling ϕ_2 from

$$\phi_2 \sim \mathcal{TN}\left(\phi_2 | \hat{\phi}_2, \omega_{22}, -1, 0\right), \tag{11}$$

a normal distribution truncated with lower bound -1 and upper bound 0. Set the proposed modulus to $m_p = \sqrt{-\phi_2}$. Proposing a frequency, we want to propose either a special or arbitrary frequency. This is achieved by introducing a discrete distribution over all possible frequency states; in this case just two. While the probabilities may be fixed at any values, the mixing of the chain is improved if a more elaborate strategy is used. Calculate a proposal probability for the special frequency as

$$p_s = \frac{w^c_{s,\omega^*} \exp\left\{-\frac{1}{2}\sigma_1^{-2}\left(\mu_1 - 2\sqrt{-\phi_2}a^*\right)^2\right\}}{w^c_{s,\omega^*} \exp\left\{-\frac{1}{2}\sigma_1^{-2}\left(\mu_1 - 2\sqrt{-\phi_2}a^*\right)^2\right\} + w^c_{s,\omega}\frac{\kappa}{\bar{a}+1}\left(2\pi\sigma_1^2\right)^{1/2}}, \tag{12}$$

where

$$\begin{aligned}
\mu_1 &= \hat{\phi}_1 + \omega_{12}\omega_{22}^{-1}\left(\phi_2 - \hat{\phi}_2\right), \\
\sigma_1^2 &= \omega_{11} - \omega_{12}\omega_{22}^{-1}\omega_{21}, \sigma_2^2 = \omega_{22} \\
\kappa &= \Phi\left(\frac{2\sqrt{-\phi_2}\bar{a} - \mu_1}{\sigma_1}\right) - \Phi\left(\frac{2\sqrt{-\phi_2} - \mu_1}{\sigma_1}\right).
\end{aligned}$$

Then, with probability p_s set $a_p = a_s$ and with probability $1 - p_s$ sample ϕ_1 from

$$\phi_1 \sim \mathcal{TN}\left(\phi_1 | m_1, \sigma_1^2, -2\sqrt{-\phi_2}, 2\sqrt{-\phi_2}\bar{a}\right), \tag{13}$$

to propose either $(m_p, a_p) = \left(\sqrt{-\phi_2}, a_s\right)$ or $\left(\sqrt{-\phi_2}, \phi_1/2\sqrt{-\phi_2}\right)$.

To calculate the relevant acceptance probability, and sample the next state, define the current and proposed root contribution to the lag polynomial and filter Z to get respective residuals,

$$E_c = \left(1 - 2m_c a_c L + m_c^2 L^2\right) Z$$
$$E_p = \left(1 - 2m_p a_p L + m_p^2 L^2\right) Z$$

Then, if a stationary pair with arbitrary frequency is proposed, the acceptance probability is calculated with

$$\alpha_{M1} = \min\left\{1, \exp\left[-\frac{1}{2}\sigma_\varepsilon^{-2}\left(E_p' E_p - Z'Z\right)\right]\kappa\right\},$$
$$\kappa = \frac{w_{s,\omega}^c}{w_0^c}\frac{3}{4\left(\bar{a}+1\right)}\frac{j_{sn}}{j_{ns}\left(1-p_s\right)f_1\left(\phi_1\right)f_2\left(\phi_2\right)}.$$

where f_1 and f_2 the truncated normal distributions in (11) and (13) respectively. If instead a special frequency is proposed the acceptance probability is

$$\alpha_{M1} = \min\left\{1, \exp\left[-\frac{1}{2}\sigma_\varepsilon^{-2}\left(E_p' E_p - Z'Z\right)\right]\kappa\right\},$$
$$\kappa = \frac{w_{s,\omega^*}^c}{w_0^c}\frac{j_{sn}}{j_{ns} p_s f_2\left(\phi_2\right)}.$$

For the reciprocal move attempting to delete a stationary root, the proposal does not involve any sampling. Instead, the balancing move treats $\phi_1 = -m_c^2$ as if it had been sampled from (11) , and depending on the current type of frequency $\phi_2 = 2m_c a_c$ from the appropriate part of (12)-(13) . The correct acceptance probabilities are then simply the inverse of the relevant expression for α_{M1} after the obvious substitutions.

Move M2 (Transitions between the stationary roots and unit roots)

Assume the root is currently stationary. In an attempt to move the pair to the boundary of the stationary region the proposal is simply $(1, a_c)$ and leaves the frequency unchanged. It does not involve any sampling and is trivial to extend to an arbitrary number of special frequencies. When balancing, the current modulus is treated as if it had been sampled from the truncated normal in (11). Computing the appropriate residuals E_c and E_p, if the current root is stationary with an arbitrary frequency, the acceptance probability is

$$\alpha_{M2} = \min\left\{1, \exp\left[-\frac{1}{2}\sigma_\varepsilon^{-2}\left(E_p' E_p - E_c' E_c\right)\right]\kappa\right\},$$
$$\kappa = \frac{w_{1,\omega}^c}{w_{s,\omega}^c}\frac{2}{3m_c}\frac{j_{us} f_2\left(-m_c^2\right)}{j_{su}}.$$

If the root currently has a special frequency, the transition to a unit root at the special frequency is instead accepted with probability

$$\alpha_{M2} = \min\left\{1, \exp\left[-\frac{1}{2}\sigma_{\varepsilon}^{-2}\left(E_p'E_p - E_c'E_c\right)\right]\kappa\right\},$$
$$\kappa = \frac{w_{1,\omega^*}^c}{w_{s,\omega^*}^c}\frac{j_{us}f_2\left(-m_c^2\right)}{j_{su}}.$$

If instead the root is currently at the unit circle, a modulus is sampled from the truncated normal in (11). The balancing move involves no sampling. Depending on the status of the current frequency the acceptance probability is the inverse of the appropriate version of α_{M2} after substituting the current value of the modulus $m_c = 1$ with the proposed value $m_p = \sqrt{-\phi_2}$.

Move M3 (Updating stationary roots)

The first local move explores the stationary region and is only available to roots that are currently stationary. A proposal is generated using the same procedure as in Move M1 and the balancing move treats the current value as if it had been sampled with the same procedure. The acceptance probability depends on the status of the current and proposed frequency. If both the current and proposed pairs are stationary with arbitrary frequency the acceptance probability is

$$\alpha_{M3} = \min\left\{1, \exp\left[-\frac{1}{2}\sigma_{\varepsilon}^{-2}\left(E_p'E_p - E_c'E_c\right)\right]\kappa\right\},$$
$$\kappa = \frac{\left(1-p_s^c\right)f_{1c}\left(2m_ca_c\right)f_2\left(-m_c^2\right)}{\left(1-p_s^p\right)f_{1p}\left(\phi_1\right)f_2\left(\phi_2\right)}.$$

where p_s^c and p_s^p are the probability of a special frequency (12) for the current and proposed root respectively. If the current root has an arbitrary frequency and a special frequency is proposed, the acceptance probability is

$$\alpha_{M3} = \min\left\{1, \exp\left[-\frac{1}{2}\sigma_{\varepsilon}^{-2}\left(E_p'E_p - E_c'E_c\right)\right]\kappa\right\},$$
$$\kappa = \frac{w_{s,\omega^*}^c 4\left(1-p_s^c\right)f_{1c}\left(2m_ca_c\right)f_2\left(-m_c^2\right)}{w_{s,\omega}^c 3\left(\bar{a}+1\right)p_s^p f_2\left(\phi_2\right)}.$$

If both the current and proposed pairs are stationary with special frequency the acceptance probability is just

$$\alpha_{M3} = \min\left\{1, \exp\left[-\frac{1}{2}\sigma_{\varepsilon}^{-2}\left(E_p'E_p - E_c'E_c\right)\right]\kappa\right\},$$
$$\kappa = \frac{w_{s,\omega^*}^c}{w_{s,\omega^*}^c}\frac{p_s^c f_2\left(-m_c^2\right)}{p_s^p f\left(\phi_2\right)}.$$

Move M4 (Updating the frequency of unit roots)

The second local move explores the frequencies of unit roots. Keeping the modulus constant, a new frequency can be sampled with a Gibbs step, similar to that used for updating real roots. This move is executed using (12) and (13) with $\phi_2 = -1$ and setting $a_p = \phi_1/2$.

4.4 Updating the innovation variance

With the selected model structure, the treatment of the innovation variance may be performed in a standard fashion. Having completed a full cycle of model selection, and updating of trend and lag polynomials compute the observed residuals,

$$E = \Psi^c(L)\left(\widetilde{Y} - X_c\beta_c\right), \quad \widetilde{Y} = (1-L)^{d_c} Y.$$

Assuming normally distributed errors, the conditional likelihood of σ_ε^2 has an inverse gamma kernel. With the conjugated inverse gamma prior on the precision this leads to

$$\sigma_\varepsilon^{-2} \big| \, Y, \xi_c \backslash \sigma_\varepsilon^2 \sim \mathcal{G}amma\left(\sigma_\varepsilon^2 | \alpha_0 + T/2, \beta_0 + E'E/2\right),$$

where T the number of observations.

4.5 Sampling latent initial values

The suggested approach circumvents the initial value problem by simply modelling the initial values as latent variables. As we may for the purpose of updating them condition on all else, the initial values are generated in three simple steps.

First, conditional on the differencing order, the trend and the lag polynomial, define

$$\upsilon_t = (1-L)^{d_c} y_t - \tau_t^{'}\beta_c,$$

for $t = d_c + 1, \ldots, T$. Using $\Psi^c(L)\,\upsilon_s = \varepsilon_s$ and reversing the time arrow, back-cast the required number of initial values υ_s^0 for $s = d_c, d_c - 1, \ldots, -\overline{r} - 2\overline{c} - \overline{d} + d_c + 1$ iteratively while sampling an innovation, conditional on the variance. Note how this is possible even when the polynomial includes unit roots, see Huerta and West (1999*a*) for a proof. Finally, extract initial values on the original scale using

$$(1-F)^{d_c} y_s = (-1)^{d_c}\left(\tau_{s+d_c}'\beta_c + \upsilon_{s+d_c}^0\right)$$

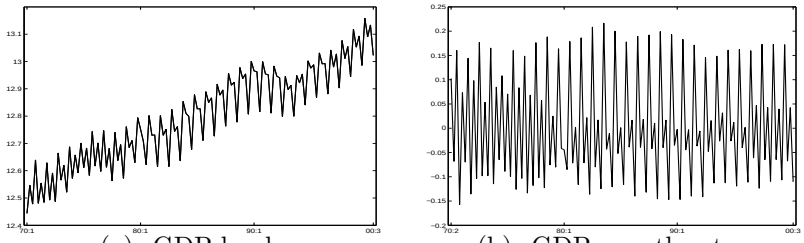
for $s = 0, \ldots, -\overline{r} - 2\overline{c} - \overline{d} + 1$.

5 The Swedish GDP

Figure 1 shows the log of Swedish real GDP, in levels as well as after taking a first difference. The quarterly data span the period from the first quarter in 1970 to the third quarter in 2000. The data is dominated by two features, a strong seasonal pattern which appears to be stable from the beginning of the eighties and a fall in GDP in the early nineties.

We specify the prior parameters as follows. For the trend component we set the maximum differencing order to $\overline{d} = 2$ and the maximum degree of the trend polynomial to $\overline{g} = 3$, i.e. a quadratic trend, with a uniform prior over all possible combinations. The scale factor in the g-prior for the parameters in the deterministic trend is set to $g = 1/T$, i.e. the prior information corresponds to the average information in one observation. We allow for a maximum of $\overline{r} = 8$ real roots and use the proposed default choice for the weights in the mixture prior, that is $w_0^r = 1/2$ and $w_s^r = w_{-1}^r = 1/4$. The expected number of active real roots is four with two stationary roots and and two negative unit roots. The uniform prior for stationary roots is truncated above at 0.99, that is $\delta = 0.01$.

Figure 1 The log of Swedish real GDP, quarterly from 1970:1 to 2000:3; levels (left) and differenced (right).



(a): GDP levels. (b): GDP growth rates.

For the complex roots we set $\overline{c} = 4$ so the real and complex roots can contribute in equal amounts to the effective lag length. The weights in the mixture prior are set to $w_0^c = 1/2$ $w_{s,\omega^*}^c = w_{s,\omega}^c = w_{1,\omega^*}^c = w_{1,\omega}^c = 1/8$. That is, we expect two complex roots with equal probability of these stationary or on the unit circle and have a seasonal or arbitrary frequency. The precision hyperparameters are calibrated using results based on available data from 1950 to 1970.

We obtain the posterior distribution from one run of the Markov chain with 250 000 replicates after having discarded 5000 replicates as burn in. The chain was monitored for convergence by studying the sample path of the model indexes, running means and the Geweke (1992) z-statistic for the constant term and error variance. To reduce autocorrelation and simplify postprocessing the chain was thinned by only retaining every fifth replicate.

Simple univariate measures of the posterior probabilities of different model configurations are shown in Table 1. It is evident that the dominant model is one with a constant term, one positive and one negative unit root and one seasonal unit root. The seasonal difference $(1 - L^4)$ that might be expected from viewing the data is thus completed. With probability 0.24 a second complex root, a stationary root with arbitrary or seasonal frequency or possibly a unit root with arbitrary frequency, is present in the model. The posterior distribution of these roots are displayed in Figure 2. It is clear that they all, to some extent mimic a seasonal unit root. This is, in particular the case for the unit root at arbitrary frequency and the stationary seasonal root. The arbitrary unit root only appears when the seasonal unit root is absent and all the probability mass above a modulus of 0.9 for the stationary seasonal root correspond to replicates when the seasonal unit root is absent. The frequency for the arbitrary stationary root has a mode close to $\pi/4$ and this root appear to be aliasing for the seasonal frequency $\pi/2$. The posterior probability of a third complex root is only 0.033 and there is no evidence of a business cycle in the data.

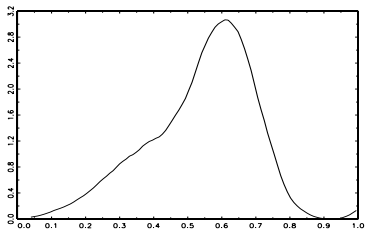
Turning to the real roots, the posterior distribution of the largest negative root, the smallest root and the largest positive root are displayed in Figure 3. For the negative root we, again, see a tendency for this root to proxy for the negative unit root when this is absent from the model. The posterior for the positive real root is well behaved does not show any adverse affects of the truncation at 0.99 enforced by the prior distribution. The posterior for the smallest root is a mixture over models with varying number of real roots and the posterior mode close to zero correspond to models with a large number of real roots.

Figure 3 also shows the posterior distribution of the constant term in models with

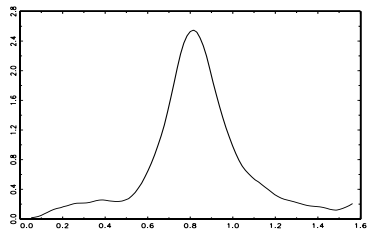
Table 1 Univariate posterior probabilities for model configurations

Number of roots	0	1	2	3	4	5	6
Real roots	0.000	0.001	0.032	0.526	0.307	0.111	0.021
Stationary real	0.001	0.032	0.520	0.310	0.113	0.022	0.002
Neg. unit root	0.010	0.990					
Complex roots	0.000	0.723	0.242	0.033	0.002		
Arbitrary stationary	0.788	0.194	0.018	0.000			
Seasonal stationary	0.883	0.111	0.006	0.000			
Unit arbitrary	0.986	0.014	0.000	0.000			
Unit seasonal	0.054	0.946	0.000	0.000			
Stochastic and deterministic trends							
Unit roots	0.000	1.000					
Degree of trend polynomial	0.183	0.766	0.046	0.004	0.001		

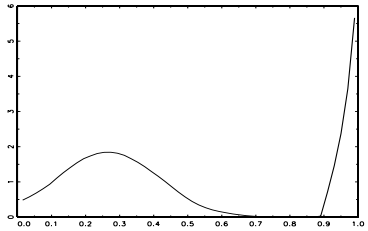
Figure 2 Posterior distribtion of complex roots.



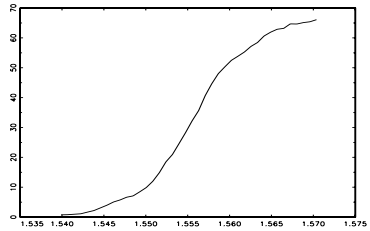
(a): Modulus of arbitrary stationary root.



(b): Frequency of arbitrary stationary root.



(c): Modulus of seasonal stationary root.



(d): Frequency of arbitrary unit root.

only a constant. The posterior mean is 0.005 corresponding to a quarterly growth rate of 0.5%. The posterior for the constant is more dispersed in models with a linear trend and has a slightly lower mean of 0.0039. The posterior mean for the trend coefficient is 0.00006 with a posterior variance of 6.8×10^{-9}, effectively removing the trend from the model.

The Markov chain explores a large range of lag length configurations as is evident from the posterior distribution for the number of real and complex roots in Table 1. Still the posterior distribution of the lag lengths is fairly concentrated with 90% of the posterior mass concentrated on lags 5 to 8 with posterior probabilities 0.408, 0.231, 0.178 and 0.097. In addition to this the unit root in the trend component is always present leading to effective lag lengths of 6 to 9.

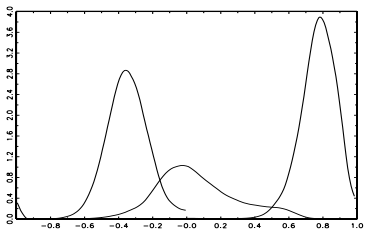
Finally, Table 2 show the posterior mean of the coefficients in the stationary polynomial $\Psi(L)$ for the combinations of number of real and complex roots with the highest posterior probabilities. The posterior means for the first five lags are relatively insensitive to the root configuration. It might thus be difficult to infer the root configuration from the AR-coefficients if these are estimated directly. The coefficients for the higher lags are quite small, this – together with the small posterior probabilities for higher lag orders – is reflected in the posterior mean for the model averaged polynomial.

6 Final remarks

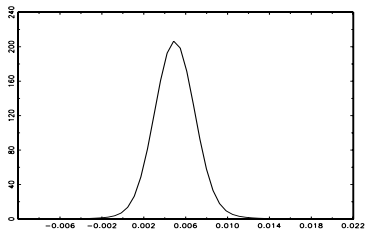
Extending the work of Huerta and West this paper demonstrates how the fundamental features of autoregressive processes, the roots of the characteristic polynomial, can be modelled directly. By focusing on the roots we are able to extract information about the dynamics of the data which is otherwise not readily available.

The modelling framework is highly flexible and automatically robustifies against misspecification by implicitly averaging over different lag lengths, number of unit roots and

Figure 3 Posterior distribution of real roots and constant term.



(a): Largest negative, smallest and largest postive root.



(b): Constant term.

Table 2 Posterior distribution of coefficients in $\Psi(L)$.

	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	p
$r = 3, c = 1$	-0.549	-0.277	-0.278	0.721	0.272		0.408
$r = 4, c = 1$	-0.556	-0.272	-0.286	0.713	0.270	-0.014	0.217
$r = 3, c = 2$	-0.559	-0.295	-0.205	0.736	0.296	0.033	0.105
Model average	-0.553	-0.278	-0.261	0.718	0.272	-0.002	

specifications for the deterministic trend. This is accomplished at the same time as inference on these issues is straightforward, the posterior probability of any given model specification can be estimated directly from the output of the Markov chain.

The flexibility comes at a cost, mainly in terms of the specification of the prior distribution. The choice of weights for the mixture prior is not always straightforward. The default choices discussed in Section 3.2 are useful starting points but can lead to counter intuitive results and should not be adopted automatically. The restriction of the prior on stationary real roots to a subset of the stationary region might also be unpalatable in some contexts. We plan to address both the specification of the prior on the components of the polynomial and alternative ways of resolving the near non-identification of the constant term in future research.

References

DeJong, D. N. and Whiteman, C. H. (1991*a*), 'Reconsidering 'trends and random walks in macroeconomic time series", *Journal of Monetary Economics* **28**, 221–254.

Dejong, D. N. and Whiteman, C. H. (1991*b*), 'The temporal stability of dividends and stock prices: Evidence from the likelihood function', *American Economic Review* **81**, 600–617.

Franses, P. H., Hoek, H. and Paap, R. (1997), 'Bayesian analysis of seasonal unit roots and seasonal mean shifts', *Journal of econometrics* **78**, 359–380.

Geweke, J. (1992), Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, *in* J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds, 'Bayesian Statistics 4', Oxford University Press, Oxford.

Green, P. J. (1995), 'Reversible jump markov chain monte carlo computation and bayesian model determination', *Biometrika* pp. 711–732.

Huerta, G. and West, M. (1999*a*), 'Bayesian inference on periodicities and component spectral structures in time series', *Journal of Time Series Analysis* pp. 401–416.

Huerta, G. and West, M. (1999*b*), 'Priors and component structures in autoregressive time series models', *Journal of the Royal Statistical Society. Series B (Methodological)* .

Hylleberg, S., Engle, R. F., Granger, C. W. J. and Yoo, B. S. (1990), 'Seasonal integration and cointegration', *Journal of Econometrics* **44**, 215–238.

Lubrano, M. (1995), 'Testing for unit roots in a bayesian framework', *Journal of Econometrics* **69**, 479–507.

Schotman, P. C. (1994), 'Priors for the AR(1) model: Parameterization issues and time series considerations', *Econometric Theory* **10**, 579–595.

Schotman, P. C. and van Dijk, H. K. (1991), 'A bayesian analysis of the unit root hypothesis in real exchange rates', *Journal of Econometrics* **49**, 195–238.