Bayes Model Averaging of Cyclical Decompositions in Economic Time Series [*]

Richard Kleijn [†]   Herman K. van Dijk [‡]

December 10, 2003

Econometric Institute Report EI2003-48

Abstract

A flexible decomposition of a time series into stochastic cycles under possible non-stationarity is specified, providing both a useful data analysis tool and a very wide model class. A Bayes procedure using Markov Chain Monte Carlo (MCMC) is introduced with a model averaging approach which explicitly deals with the uncertainty on the appropriate number of cycles. The convergence of the MCMC method is substantially accelerated through a convenient reparametrization based on a hierarchical structure of variances in a state space model. The model and corresponding inferential procedure are applied to simulated data and to economic time series like industrial production, unemployment and real exchange rates. We derive the implied posterior distributions of model parameters and some relevant functions thereof, shedding light on a wide range of key features of each economic time series.

Keywords: Fourier analysis; time series decomposition; model averaging; state space models; Markov Chain Monte Carlo

1 Introduction

The frequency domain properties of an economic time series may provide a useful complement to its time domain properties. The link between the time and frequency domain for a given time series $\{y_t\}_{t=1}^T$ can be described by the Fourier representation

$$y_t = \mu + \sum_{j=1}^{\lfloor T/2 \rfloor} r_j \cos(\lambda_j t - \phi_j), \tag{1.1}$$

with $\lambda_j = 2\pi j/T$ and $\phi_{\lfloor T/2 \rfloor} = 0$ when T is even. For a clear introduction to this subject we refer to e.g. Harvey (1993, Ch.6). When treated as a descriptive model for economic time series , equation (1.1) has μ, $r_1, \ldots, r_{\lfloor T/2 \rfloor}$ and $\phi_1, \ldots, \phi_{\lfloor T/2 \rfloor}$ as its parameters. The number of parameters equals the number of observations, hence guaranteeing a perfect fit. When the terms $r_j \cos(\lambda_j t - \phi_j)$ are rewritten as $a_j \cos \lambda_j t + b_j \sin \lambda_j t$ the regressors have nice orthogonality properties. The drawbacks of regarding equation (1.1) as a time series model are also obvious. The perfect fit prevents any useful statistical inferences to be made and more importantly, implicitly it is assumed that the time series is exactly repeating itself infinitely far into the future. This regularity may be appropriate for some time series in e.g. geology, climatology, or astronomy. However, economic time series tend to be less deterministic and more irregular. As a descriptive procedure and data analysis tool the Fourier representation and the closely related sample periodogram is helpful. However, when e.g. forecasting is the objective, a model based approach may be more helpful. By using fewer than $\lfloor T/2 \rfloor$ cycles and introducing some randomness in the cycles, we can specify a model which acknowledges that the data are not perfectly cyclical, while still some exploitable structure is provided. Our model has the Fourier representation as its limiting case, thus ensuring that even for the messiest of time series at least one of the models will fit the data. Other model-based approaches include West (1995), Huerta and West (1999), and Harvey et al. (2002). West (1995), Huerta and West (1999) provide a method using Bayesian prior specification on parameters of a (high-order) AR model based on roots of the AR polynomial. They allow for uncertainty on the number of cycles, which are themselves directly associated with complex AR roots. In their approach all cycles are driven by a single innovation term, which may be overly restrictive as it introduces some complex structure in terms of cross-spectra. Harvey et al. (2002) describe an unobserved components structural time series model with a trend, an irregular term and a single cycle, but with the advantage that each component has its own innovation term. They find prior information on the period of oscillation to be helpful in avoiding unreasonable values.

In this paper we introduce a model-based, explicit unobserved components model which allows for stochastic cycles (as opposed to deterministic cycles). In particular we allow for an *unknown* number of cycles and use Posterior Odds analysis to decide on the number of cycles

or, more appropriately, use its implied Posterior Model Probabilities as weights in a mixture of models with different numbers of cyclical components to capture the inherent uncertainty on the number of cycles. Moreover, different cycles may be driven by different innovations. We consider this an important feature, since it is an open question whether trend, business cycle and seasonal effects are really driven by one single innovation.

Our method is illustrated on simulated data. Next, several empirical applications show how the method can be used as a tool for empirical work with wide applicability. For inferential purposes we make use of a Markov Chain Monte Carlo (MCMC) simulation based Bayesian analysis. The posterior simulator for our model class is a combination of data augmentation and a Gibbs sampler, which contains a simulation smoother step, see de Jong and Shephard (1995) or similar work by Carter and Kohn (1994) and Frühwirth-Schnatter (1994), and a Metropolis-Hastings step. Our simulation based Bayesian inference is an extension of the work of Koop and van Dijk (2000), but their focus is on testing for stationarity in an unobserved components model.

A contribution of this paper is the convenient and useful reparametrization of the model which provides a structured way of specifying prior information. A second contribution is the development of an efficient candidate generator for the model parameters of unobserved cycle components. It is specifically tailored for the Metropolis-Hastings step in the sampler for this model. The simulation based approach allows us to do inference directly on interesting functions of parameters. In particular, we can derive the full posterior distributions of impulse response functions and half-lives and we can estimate spectral densities.

The outline of the paper is as follows: In Section 2 we start by introducing the model class that we are using throughout this paper. We also outline the Bayesian procedures for these models. The artificial data experiment that concludes Section 2 illustrates the use of methods in practice. In Sections 3 to 5 the cyclical decomposition procedure is applied in order to investigate the cyclical properties of three empirical time series, viz. US Industrial Production, US unemployment and the DM/USD real exchange rate. For each of the series some specific aspects, relevant to the series under consideration, which can be analyzed by the decomposition procedure is highlighted. The results include tests for constancy of the underlying mean, posterior model probabilities, posterior densities of parameters but also of some other quantities that may of interest and can be expressed as (more complicated) functions of the parameters and data. In particular we report full forecast densities, impulse responses, their implied half-lives, and spectral density estimates, which are notoriously difficult to estimate directly from the data. Section 6 concludes with a discussion of the results.

2 Cyclical Decomposition Procedure

2.1 Introduction

For time series analysis the frequency domain often provides valuable information on the dynamics of the series. Other parts of the analysis are more appropriately done in the time domain. Frequency and time domain properties are more often than not analyzed separately. In this paper we aim to provide a model for the time domain that more explicitly incorporates frequency domain properties with the underlying motivation that one domain could complement the other. In particular, our model consists of a composition of cycles, a feature it shares with the familiar discrete Fourier decomposition. This structure provides a very flexible functional form. We aim to obtain a structure that is as flexible as possible, without the danger of overfitting. We carefully try to avoid the noise fitting that would occur when fully unrestricted estimation would be tried. By expressing the model in terms of a hierarchical structure and specifying a prior accordingly we avoid these intricacies, while at the same time it facilitates the elicitation of prior information.

2.2 Specification

The core of the model consists of a composition of several (unobserved) cyclical components. We use the same functional form as e.g. Harvey and Streibel (1998). It takes the form of a (restricted) two-dimensional VAR(1) model, see equation (2.1). A univariate AR(2) component with a parameter restriction to ensure a pair of complex roots is another popular choice for a cyclical component. The VAR(1) specification gives a more regular cycle which is also reflected in a more pronounced peak in the theoretical spectrum. These stochastic cycles have a deterministic cycle as their limiting case.

Our model consists of components meant to describe long-term movements, cyclical features and irregular movements of the data. The full model is given by:

$$y_t = \mu_t + \sum_{c=1}^{C} \psi_{c,t} + \xi_t \tag{2.1a}$$

$$\mu_t = \mu_{t-1} + \eta_t \tag{2.1b}$$

$$\begin{pmatrix} \psi_{c,t} \\ \psi_{c,t}^* \end{pmatrix} = \rho_c \begin{pmatrix} \cos \omega_c \pi & \sin \omega_c \pi \\ -\sin \omega_c \pi & \cos \omega_c \pi \end{pmatrix} \begin{pmatrix} \psi_{c,t-1} \\ \psi_{c,t-1}^* \end{pmatrix} + \begin{pmatrix} \kappa_{c,t} \\ \kappa_{c,t}^* \end{pmatrix}, c = 1, \ldots, C \tag{2.1c}$$

4

for which we assume Gaussian innovations:

$$\xi_t \sim N(0, \sigma_\xi^2)$$

$$\eta_t \sim N(0, \frac{6}{T}\sigma_\mu^2)$$

$$\begin{pmatrix} \kappa_c \\ \kappa_c^* \end{pmatrix}_t \sim N(0, (1 - \rho_c^2)\sigma_c^2 I_2)$$

The key features of (2.1) are summarized as follows:

- A stochastic trend μ_t, which could straightforwardly be extended with a drift term or replaced by a smoother trend specification. This term provided the local mean μ_t of the series is, see equation (2.1b). Note that σ_μ^2 here represents the expected sample variance in a sample of size T from the process $\{\mu_t\}$, which does exist in contrast with the unconditional variance.

- C damped stochastic sinusoids ψ_c with frequencies ω_c and damping factors ρ_c. The parameters ω_c are the frequencies and ρ_c are the damping factors of the damped stochastic sinusoids, see equation (2.1c)

- an irregular term ξ_t, additive noise

In a Bayesian analysis the initial conditions play a crucial role for correcting the pathological behaviour of the likelihood near the unit root, see e.g. Schotman and van Dijk (1991) on initial conditions in the context of an AR(1) model or Zivot (1994) within an unobserved components model.

We specify a multivariate normal distribution for the initial values of the states, with its mean and covariance matrix chosen to match the unconditional moments as recommended in e.g. Harvey (1989).

$$\begin{pmatrix} \psi_{c,0} \\ \psi_{c,0}^* \end{pmatrix} \sim N(0, \sigma_c^2 I_2)$$

We assume μ_0 to be fully diffuse, such that it acts as an intercept for the entire model.

The model essentially consists of the sum of an ARIMA(1,1,0), several ARMA(2,1) components and a white noise term, which results in a high-order ARIMA model with complicated parameter restrictions when represented in single-equation form. This indicates that, as argued in Section 2.1, we have a very flexible time series model.

2.3 Parametrization

In a model with potentially many components it is particularly important that the components can be clearly distinguished and are not overlapping. If this is not the case, the components

may spuriously pick up other frequencies and hence hamper identification of the individual components.

Since we are dealing with potentially non-stationary data, the unconditional variance of the trend component does not exist for $T \to \infty$. We propose to compare the unconditional variance of the cycles with the expected sample variance of the random walk over the sample period. The latter number is finite and hence is a metric directly comparable to the variances of the cyclical components. This corresponds to the intuition that with a longer span of data cycles with larger periods can be identified.

The expectation of the overall sample variance σ^2 is divided between the non-stationary components (fraction ν) and stationary part. This can be described as an hierarchical structure on the variances. We therefore introduce a transformation from $(\sigma_\mu^2, \sigma_1^2, \ldots \sigma_C^2, \sigma_\xi^2)$ to $(\sigma^2, \nu, \lambda_1, \ldots, \lambda_C)$ which is defined by

$$\sigma_\mu^2 = \nu \sigma^2$$

$$\sigma_\psi^2 \equiv \sigma_\xi^2 + \sum_{i=1}^{C} \sigma_i^2 = (1-\nu)\sigma^2$$

$$\sigma_i^2 = \lambda_i \sigma_\psi^2$$

$$\sigma_\xi^2 = (1 - \sum_{i=1}^{C} \lambda_i)\sigma_\psi^2$$

with Jacobian

$$\left| \frac{\partial \left(\sigma_1^2 \quad \cdots \quad \sigma_C^2 \quad \sigma_\xi^2 \quad \sigma_\mu^2 \right)'}{\partial \left(\lambda_1 \quad \cdots \quad \lambda_C \quad \sigma^2 \quad \nu \right)} \right| = (1-\nu)^C (\sigma^2)^{C+1}.$$

We refer to the Appendix for details on the derivation. Thus, the Jacobian factor has a relatively simple expression and allows us to operate in the transformed space in a relatively easy way.

2.4 Prior

This parametrization simplifies eliciting priors, both informative and noninformative. An example of a weakly informative prior (that is, relative with respect to the information in the likelihood) is obtained when we specify for the overall variance a diffuse prior on its log. The variance of the cyclical components is given as a fraction of the overall variance. Dirichlet

priors are the standard choice for the fraction parameters.

$$\nu \sim B(\frac{1}{2}, \frac{1}{2})$$
$$\lambda_i \sim D(1, \ldots, 1)$$
$$\omega_i \sim U(0 \leq \omega_1 \leq \cdots \leq \omega_C \leq 1)$$
$$\rho_i | \omega_i \sim B(2/\omega_i, 1)$$

Here the prior on ω_i is uniform on a triangular subset to ensure identification of the cycles by ordering them according to frequency. This could alternatively be written as standard Dirichlet on the increments: $(\omega_1, \omega_2 - \omega_1, \ldots, \omega_C - \omega_{C-1}) \sim D(1, \ldots, 1)$. The parameters of the Dirichlet distributions could also be made dependent on the number of cycles to ensure that the frequencies are centered closer around an equidistant grid when the number of cycles increases. In that case the prior can be specified as $(\omega_1, \omega_2 - \omega_1, \ldots, \omega_C - \omega_{C-1}) \sim D(\theta_C, \ldots, \theta_C)$, with θ_C a parameter that depends on the number of cycles in the model. The closer centering could be accomplished by using $\theta_C = C \frac{\lfloor T/2 \rfloor - 1}{\lfloor T/2 \rfloor - C}$. For $C \to \lfloor T/2 \rfloor$ the parameter θ_C goes to infinity. This reflects that in the limiting case of $T/2$ cycles the distribution of the frequencies should degenerate to the fixed Fourier frequencies without any spread around them. This improves identification of the cycles. For small values of C, θ_C is approximately equal to C.

The period of the ith cycle is $2/\omega_i$. In each observation period only a fraction ρ_i of the information on amplitude and phase of the ith cycle is carried forward. After a complete cycle only a fraction ρ^{2/ω_i} of the initial information remains. The value ρ^{2/ω_i} can be thought of as a measure of smoothness of the cycle. In contrast to ρ it is comparable between cycles of differing frequency. The interpretation in terms of smoothness of the cycles is helpful for prior elicitation. We have chosen to specify a uniform prior on smoothness. This implies the mentioned beta-distributions for the damping parameters ρ_i.

The prior we specified is improper. This improperness is strictly due to $p(\sigma^2)$. This poses no problem for Posterior Odds analysis as this (nuisance) parameter occurs in all models throughout the entire model class with exactly the same interpretation and prior distribution. Hence, it cancels in the ratio of marginal likelihoods. It serves merely as a scale parameter and has similar orthogonality properties as the variance parameter in the linear regression case and therefore our procedure is scale invariant.

Additionally, one can specify prior information on the number of cycles C. Straightforward examples are e.g. equal probabilities, (truncated) geometric or Poisson. This prior acts as a tuning parameter to influence the tendency of overfitting. This matter will be discussed in more detail in Section 2.7.

2.5 Posterior Simulator

We are primarily interested in conducting posterior inference on (functions of) the model parameters. For that purpose we need the posterior density $p(\theta|Data)$ of the model parameters $\theta = (\{\lambda_i, \omega_i, \rho_i\}, \sigma^2, \nu)$. Direct Bayesian analysis of $p(\theta|Data)$ is analytically untractable, but we can apply a technique known as data augmentation, see Tanner and Wong (1987). We extend the parameter space by treating the unobserved states $\alpha_t = (\{\Psi_i, \Psi_i^*\}, \mu)$ as extra parameters and the resulting joint posterior $p(\theta, \alpha|Data)$ is much easier to simulate from.

We sample from the extended model and by marginalizing with respect to the additional parameters we can easily translate results back to the original model, i.e.

$$p(\theta|Data) = \int p(\theta, \alpha|Data)d\alpha.$$

In the remainder of this section we explain our simulation algorithm which is the basis of our posterior inference. For a recent general survey of the use of simulation methods in Bayesian analysis, see Geweke (1999). For $p(\theta, \alpha|Data)$ we use a Gibbs sampler, see e.g. Gelfand and Smith (1990) or Casella and George (1992). One iteration of the Gibbs sampler for the joint distribution of several random variables consists of going through the sequence of drawing each of the random variables conditional on the most recently obtained value of the remaining variables. The algorithm consists of drawing from the following conditional posterior densities:

If we denote the vector of all model parameters by $\theta = (\{\sigma_i^2, \omega_i, \rho_i\}, \sigma_\xi^2, \sigma_\eta^2)$ and the set of unobserved states in each iteration by $\alpha = (\{\Psi_i, \Psi_i^*\}, \mu)$, then the sampler can be described by

1. $p(\alpha|\theta, y)$

2. $p(\theta|\alpha, y)$

The first step is implemented as a simulation smoother. We employ the version by de Jong and Shephard (1995). The implementation of the second step is a more intricate matter.

An inherent difficulty with data-augmentation is that the strong dependence of sampled model parameters on the state vector, and the other way around, introduces strong correlations in the Markov chain and this leads to relatively slow mixing. Therefore, it is essential to sample model parameters conditionally on the states as efficient as possible in the sense that it should not introduce much additional correlation. Hence, it is desirable to sample all model parameters in one single Gibbs step. Direct sampling of all model parameters simultaneously is not feasible here, but we have developed an efficient, specifically tailored Metropolis-Hastings step which comes close. It is based on a candidate density which is both convenient for sampling and closely approximates the true conditional posterior density. We refer to the appendix for technical details.

2.6 Model Choice and Model Averaging

Model choice and model averaging are common concepts in Bayesian econometrics. Both are based on posterior odds analysis. In our case we restrict the model class to include all models up to some specified number of cycles. Half the sample size is a natural upperbound arising from the standard (deterministic) harmonic decomposition. For our purposes a maximum of 4 or 5 cycles seems reasonable. Otherwise we would merely be fitting noise with the extra cycles. We would like to note that our method is quite robust for such an overfitting problem as the models with a high number of cycles are given low posterior model probability. An intuitive explanation for this phenomenon is that the parameter uncertainty in that case starts to outweigh the 'certainty' introduced by the better fit. Section 2.7 provides an illustration of this claim.

If one assigns equal prior probabilities to all models in the class of models taken into consideration, the resulting ratio of the posterior probabilities of the models is the Bayes Factor. We calculate the Bayes Factors as the ratio of marginal likelihoods. The marginal likelihoods are obtained from the posterior sample using the harmonic mean estimator

$$
\begin{aligned}
\frac{1}{\hat{m}(y)} &= \frac{1}{n} \sum_{i=1}^{n} \frac{w(\theta_i)}{L(y|\theta_i) p(\theta_i)} \\
&= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{L(y|\theta_i)}
\end{aligned}
\tag{2.2}
$$

It is often preferable to work with the logarithm of the marginal likelihood. Equation (2.2) is not very attractive as it involves the summation of the inverses of likelihood values, which potentially differ by many orders of magnitude. A numerically stable approach is obtained when we specify the procedure in terms of log-likelihoods, which we denote by $LL_i = \log L(y|\theta_i)$ and their minimum by $\overline{LL}$.

$$
\begin{aligned}
\log \hat{m}(y) &= -\log \frac{1}{n} \sum_{i=1}^{n} \frac{1}{L(y|\theta_i)} \\
&= \overline{LL} - \log \frac{1}{n} \sum_{i=1}^{n} \exp(\overline{LL} - LL_i)
\end{aligned}
$$

We note that the method by Chib (1995), which uses the output of a Gibbs sampler efficiently, cannot be applied here because the constants of integration are not known for all conditional densities. Also, the Laplace approximation cannot be used as it does not take into account multimodality of the posterior distributions, which can occur in this model.
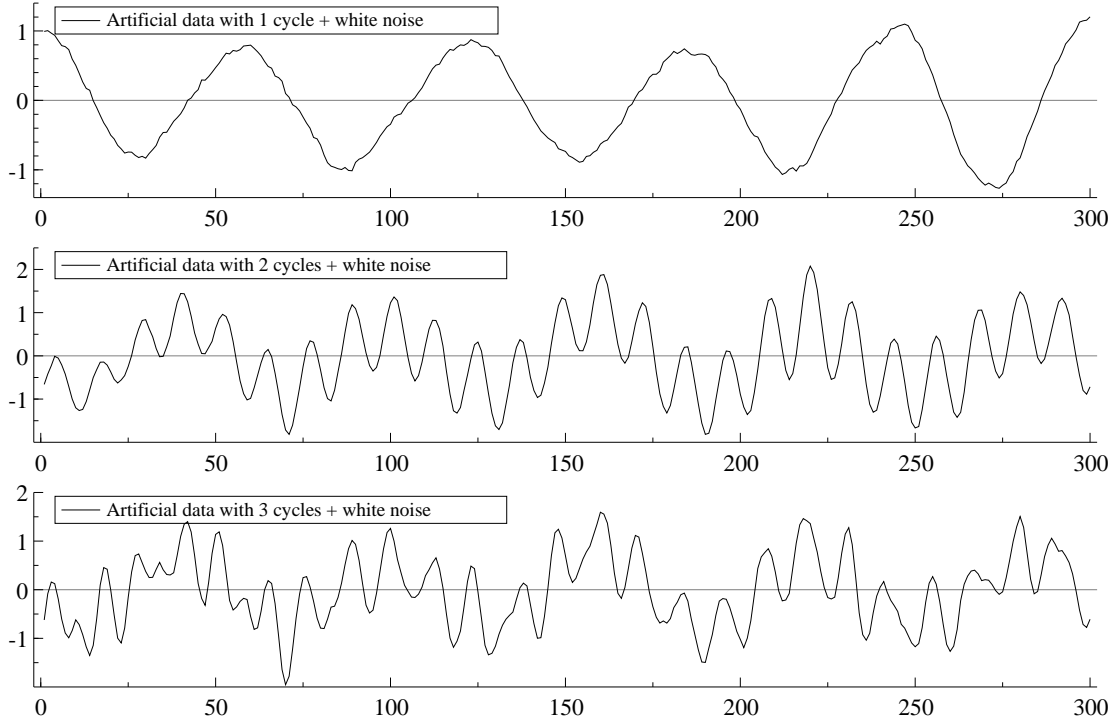
Figure 1: Artificial data with high ρ and small σ_ξ^2.

2.7 Artificial Data Experiments

We illustrate the possibilities and necessity of the procedure with two artificial datasets. Each provides an important insight. The first experiment is based on simulated data with very smooth cycles. The second data set is based on much less regular cycles.

2.7.1 Artificial data with smooth cycles

We have simulated 3 series of 300 observations. The first series has 1 cyclical component with a period of 60 periods. The second series additionally contains a cycle with 12 periods and the third data set adds an 8-period cycle. The series contain no trend component. The damping factors are chosen close, but not equal, to 1. The overall variance σ^2 is equal to 1. The variance shares λ_c of the cycles are chosen such that the variance of the first cycle is twice as large as the variance of the second cycle and four times the variance of the third cycle. The irregular term receives a very small share of the variance. The generated series are shown in Figure 1.

For each of the 3 series, we estimate 5 models, in which the number of cycles ranges from 1 to 5, using the MCMC procedure. From the posterior sample we have calculated the marginal

Table 1: Posterior model probabilities for artificial data example.

simulated model	#cycles in estimated model				
#cycles	1	2	3	4	5
1	0.531	0.252	0.112	0.077	0.028
2	0.000	0.454	0.320	0.179	0.046
3	0.000	0.000	0.516	0.344	0.140

Note: Posterior probabilities under equal prior probabilities.

likelihoods and resulting model probabilities, based on equal prior model probabilities of 1/5. These are shown in Table 1.

For each data set, the 'true' model gets the highest posterior probability. Models with 'too many' cycles get less and less posterior probability when the number of cycles is increased. Thus, the tendency for overfitting is limited. More importantly, models which contain too little cycles are firmly rejected, thereby excluding potentially strange outcomes, resulting from a model that is not flexible enough to accommodate the features of the data, to enter the final posterior results after weighing by posterior model probability.

2.7.2 Artificial data with less regular cycles

Next, we simulated a series of 300 observations with 2 cycles of 120 and 12 periods, respectively with smaller damping factors. Also, the irregular term accounts for 40% of the total variation of the series. The result is a more irregular cyclical pattern. We calculated model probabilities as before. We now find that the model probabilities give a much less clear-cut indication on the 'correct' number of cycles, although the model with only 1 cycle is still firmly excluded by the procedure. Apparently, the data is not informative enough to choose a single model with specific number of cycles.

A more pronounced posterior statement on the minimal number of cycles required for an adequate description of the dynamics in the data would naturally be obtained by attaching prior probabilities to the number of cycles. A researcher might for instance believe that mechanisms that could generate cyclical behavior in economic time series plausibly account for 1, 2 or 3 distinguishable cycles, whereas 9 or 10 cycles would be mere noise fitting. A straightforward prior that can describe such information is a geometrically declining sequence of prior model probabilities, up to some pre-specified maximum number of cycles, for instance

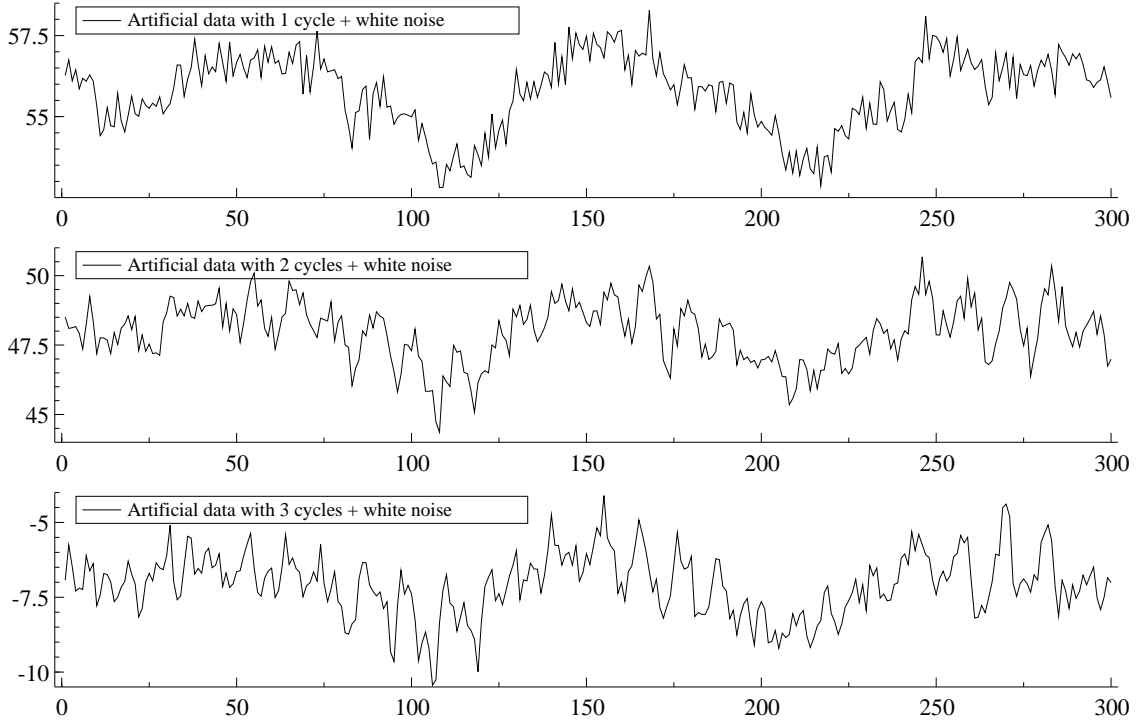$$Pr(C = c) \propto \left(\frac{1}{2}\right)^c, \quad \text{for } c = 1, \dots C_{\max}$$

11

Figure 2: Artificial data with moderate ρ and σ_ξ^2.

Such a prior would downweigh models with a higher number of cycles but it will be unable to fully compensate the low marginal likelihoods of the models with too few cycles. It effectively favours the smallest model that adequately describes the data. Note that the number of cycles should never exceed $\lfloor T/2 \rfloor$, because for that number of cycles the Fourier decomposition would guarantee a perfect fit.

Although such a prior would put less probability mass on larger models, they will still receive some weight. Thus, the question remains how a model with more cycles than strictly necessary would affect posterior inference. For the purpose of answering this, we investigate in more detail the case where the generated data contains 2 cycles and the estimated model has 4 cycles, 2 of which are actually superfluous but which catch some (non-smooth) coincidental cycles that arise in small samples of noise (the sample size here is 300).

In the posterior densities of the frequency parameters ω_c in Figure 3 we observe an interesting phenomenon. The cycle of the lowest frequencies that is present in the data is sometimes picked up by the first cycle of the estimated model and sometimes by the second one. The higher frequency cycle is picked up by either the second or third cycle in the estimated model. The fourth cycle in the model seems to pick up some spurious cycles in a wide range of frequencies, consistent with the flat spectrum of white noise. The third cycle
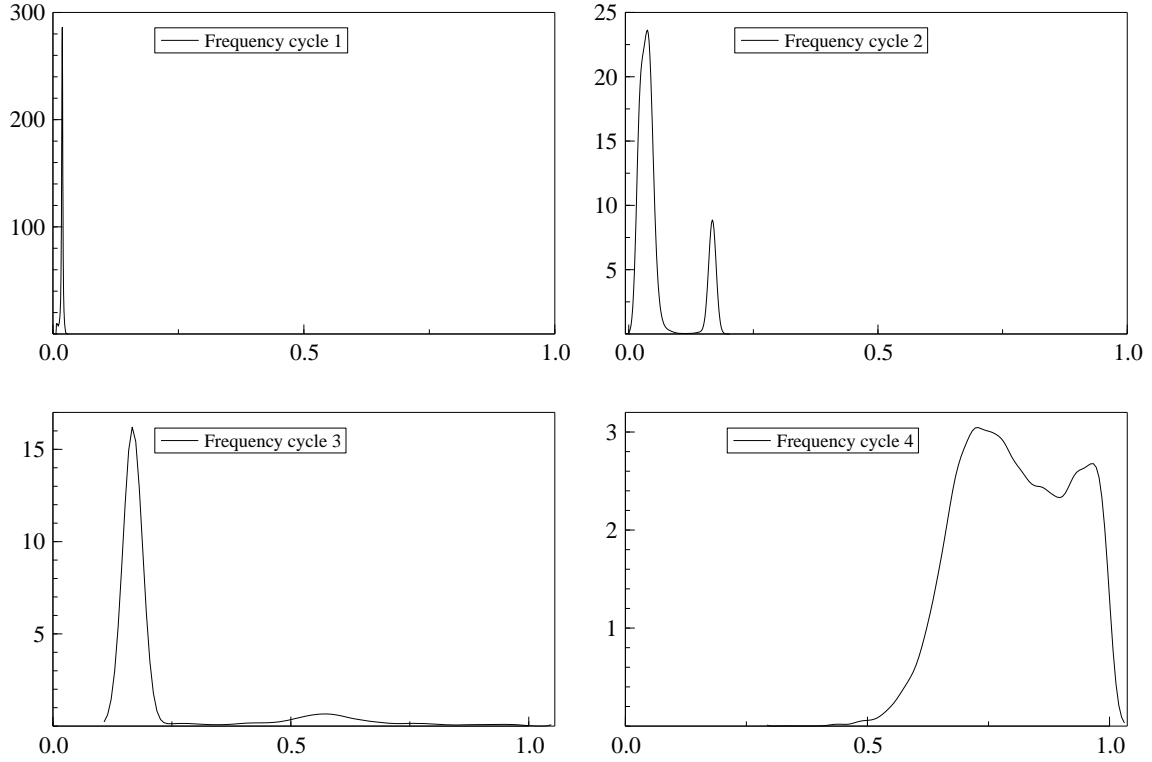
12

Figure 3: Bimodality in posterior density of frequency parameters. Simulated data contains 2 cycles, the estimated model has 4 cycles.

seems to alternate between the higher frequency cycle and noise fitting. A few observations can be based on the resulting multi-modal posterior densities. First, had one used Maximum Likelihood to estimate the model, then only one of the modes would be the MLE. The uncertainty of the asymptotic curvature-based variance estimators of the MLE would not take into account the existence of the other modes. These problems become even more apparent for the cycles of higher frequency, where the likelihood becomes very 'hairy' picking up cycles only at the frequencies that coincidentally match the noise in the data. Apart from that, any gradient-based optimization routine would become extremely sensitive to the starting values.

A second observation that can be made on the basis of the multi-modal densities concerns the problems that high-correlation MCMC chains will face. In our procedure, the sampler mixes slowly because the data augmentation inevitably introduces dependence between the 2 Gibbs steps. However, our candidate draw for the model parameters depends on the previous parameter draw only through the simulated states. And these states strongly depend on the observed data. If one uses for example a random walk Metropolis step or any other candidate that directly depends on the previous parameter draw, one essentially confines the sampling process to a localized search. In this way it becomes far more likely that other modes of the

13

posterior density will be entirely missed.

So, although the posterior densities may look strange, they more accurately describe the inherent uncertainty on the number and frequency of the cycles than the MLE would or what a more correlated MCMC procedure would indicate. The researcher is seldom primarily interested in just the posterior distribution of the model parameters, but in more interesting functions of parameters and data. This model class involves cycles and hence it is natural to investigate some of its frequency domain properties. In particular we try to estimate the frequency spectrum and see how it is influenced by the number of cycles in the model. In the subsequent sections many more possible quantities of interest are reported.

Since we generated these data ourselves, we can derive the theoretical frequency spectrum from the parameters of the process. It is a standard result that the spectrum at frequencies $\lambda \in [0, \pi]$ is given by

$$(2\pi)^{-1} \frac{1 + \rho^2 - 2\rho \cos \omega \pi \cos \lambda}{1 + \rho^4 + 4\rho^2 \cos^2 \omega \pi - 4\rho(1 + \rho^2) \cos \omega \pi \cos \lambda + 2\rho^2 \cos 2\lambda} \sigma_\kappa^2$$

for the cycle component (2.1c) and by $(2\pi)^{-1}\sigma_\xi^2$ for the irregular term, see e.g. Harvey (1989, Ch. 2). In the model the disturbances are assumed to be uncorrelated over the components. Therefore, all cross-spectra are zero and the spectrum of $\{y_t\}$ can simply be computed as the sum of the spectra of its components. For the case of 2 cycles, this is shown in the first panel of Figure 4 on the top row. The theoretical spectrum clearly shows 2 peaks corresponding to the frequencies of the 2 cycles. The second panel displays the sample spectrum which is calculated from the sample autocorrelations of the simulated data. It is a well-known result that the accuracy of the estimated sample spectrum does not decrease with sample size. This accounts for the hairy behavior of the sample spectrum. The top right panel shows a smoothed version of the sample spectrum (based on a Parzen window) to reduce the hairiness while retaining the main spikes. The resulting graph is still not fully satisfactory as it gives the impression that there are 2 peaks around the frequency of the second cycle. Alternatively, one could calculate a parametric estimate of the spectrum. A common approach consists of fitting an ARMA model to the data and calculating its induced spectrum.

The results of a parametric estimate based on our cycle decomposition procedure are reported on the middle and bottom rows of panels in Figure 4. For each draw from the posterior density we have calculated the induced power spectrum. At each frequency the posterior sample median of the spectrum is reported, but the full distribution of the spectrum at each frequency can be estimated from the sample output. In this way we can fully account for the parameter uncertainty.

The graph clearly shows that the model with only 1 cycle produces 1 very broad peak in an effort to encompass the 2 sharp peaks from the true process. This model is clearly misspecified, but its posterior probability is smaller than 10^{-18} times the probability of the
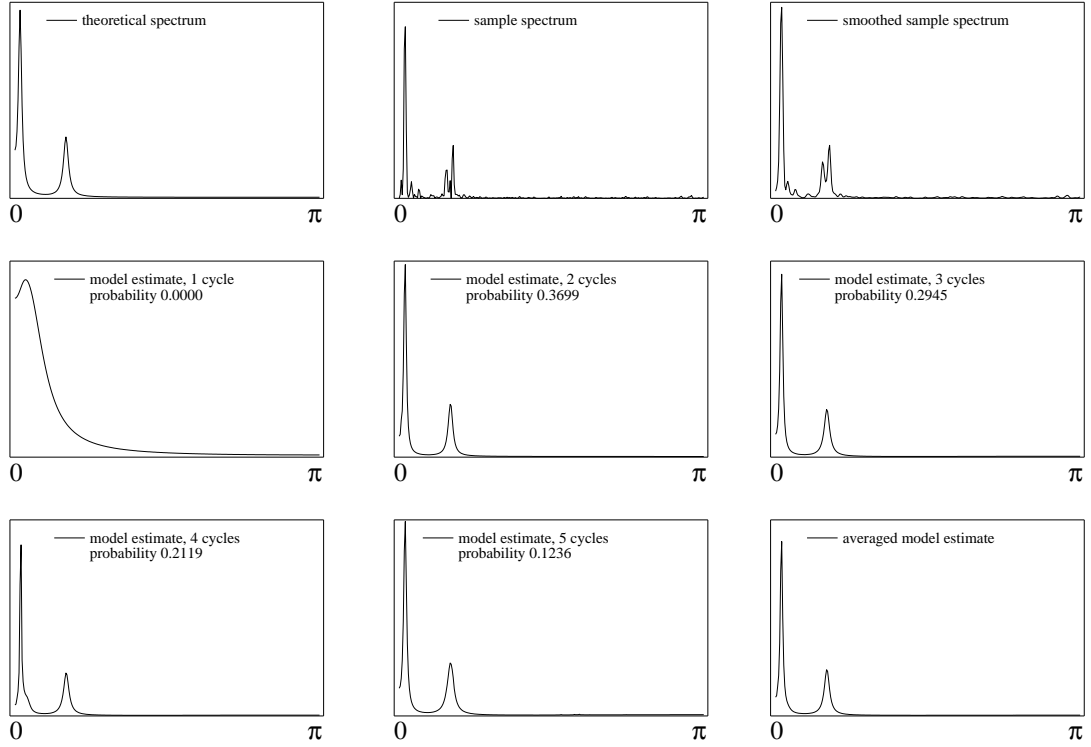
14

Figure 4: Theoretical, sample and estimated spectra for the artificial data.

model with 2 cycles. The model with 2 cycles accurately recovers the shape of the theoretical spectrum. The models with 3, 4 and 5 cycles yield very similar spectra. For a single draw from the posterior density, the superfluous cycles in these models give a third, fourth and fifth peak in the spectrum. The uncertainty of these cycles is so large that their peaks are not very sharp and they occur on different frequencies for different draws such that on average these superfluous peaks are spread out over the entire frequency range. The bottom right panel displays the estimated spectrum using a weighted average over the models using the calculated model probabilities. In this way we have not only fully accounted for the parameters uncertainty, but also for the model uncertainty within this class of cyclical models. Also, the parametric spectrum estimate recovers the theoretical spectrum more accurately than the nonparametric versions.

Summarizing, we find that a model with too few cycles is firmly rejected by our procedure. Most probability is generally attached to the 'correct' model. Model with 'too many' cycles are given some posterior probability, but this overfitting is quite harmless as the extra cycles are very uncertain and taking the uncertainty into account by averaging over the posteriors draws tends to cancel out their effect. Hence, their effect on posterior quantities tends to be negligible. Our procedure thus seems to be robust for misspecification.

3   US Industrial production

Industrial Production is a prominent cyclical sector of the economy. Therefore, Industrial Production seems a suitable series to investigate using our decomposition procedure. The questions that we seek to answer are the following:

- What is/are the durations of the growth cycles?

- Has the average growth rate changed over time?

- Can we produce a forecast incorporating all parameter and model uncertainty?

The date that we use are monthly US Industrials Production from 1960:1 to 2003:10, obtained from Economagic. The data are transformed to a 12-month growth rates.

Table 2: Posterior model probabilities for US Industrial Production.

| | #cycles in estimated model | | | | | |
trend	1	2	3	4	5	total
no	0.000	0.000	0.005	0.052	0.035	0.092
yes	0.000	0.000	0.000	0.521	0.386	0.908

Note: Posterior probabilities under equal prior probabilities.

We have estimated the model with the MCMC procedure. The model probabilities are reported in Table 2. The models with a stochastic trend component together get 91% posterior probability, providing evidence for non-stationary behavior of the Industrial Production growth rate. At least 4 cycles are required for an adequate description of the dynamics.

The estimated spectrum of the stationary components is plotted in Figure 5. It has 2 easily recognizable peaks at frequencies 0.102, and 0.217 corresponding to periods of oscillation of 62 and 29 months. There are also 2 smaller peaks at 0.329 and 0.731, describing cycles with cycle lengths of 19 and 9 months. Of these cycles the first one is the most prominent one. It is assigned almost half of the total variance. Its 5 year period is in a range commonly regarded to be related to the business cycle. However, attaching a structural or economic interpretation to the individual cycles is not our primary focus here.

The smoothed states from the model with a stochastic trend and 4 cycles are shown in Figure 6. The stochastic trend here has the interpretation of a time-varying mean growth rate. The time-varying mean component essentially remains after removing all cyclical components, which have mean 0 by construction, from the raw growth rates. We find that it has fallen from around 2.4% per year in the early 1960-s to around 1.0% per year in 1980. The 1990-s are
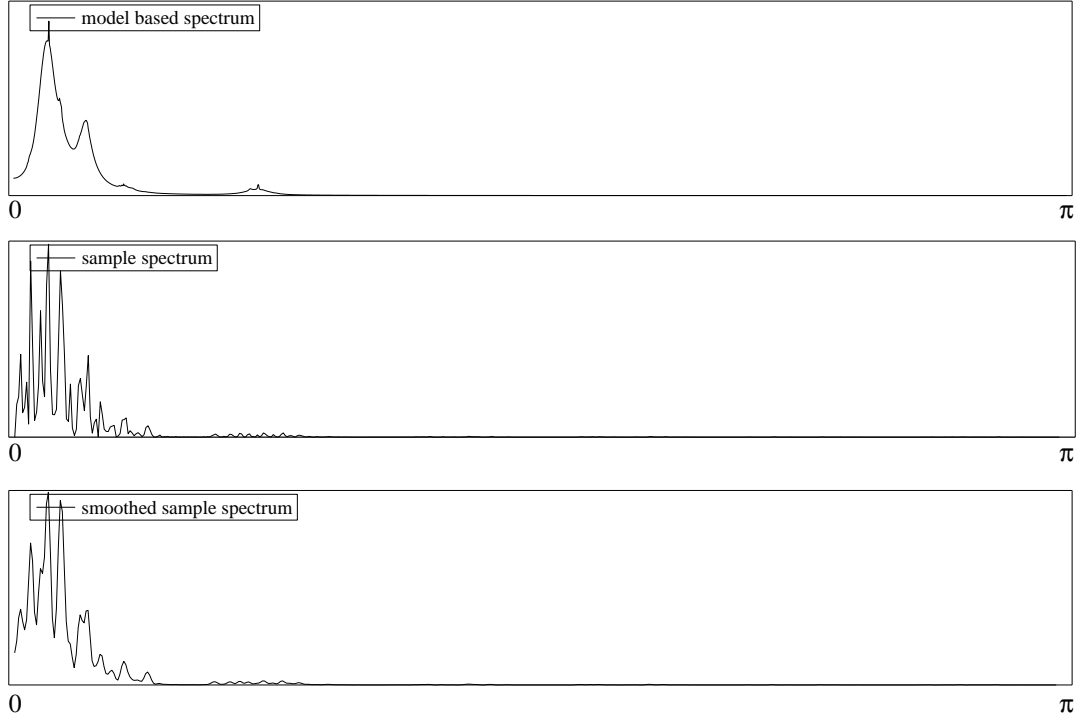
Figure 5: Sample spectra and model based spectrum of stationary components for US industrial production data.

characterized by a long period without recession during which the idea of the 'new economy' with permanent higher growth rates gained support. The higher growth rates in the 1990-s did not fit in the regular cyclical pattern, and therefore have been qualified as a permanent effect to the time-varying mean. Around 2000 this increase had already been neutralized again by subsequent changes in the mean μ_t.

The smoothed states of the first cycle show that in the 1990-s an expansion period was followed by a growth slowdown, but that it did not fully enter a contraction and it started a new expansion period in the second half of the 1990-s.

The last cycle seems to capture some of the seasonality in the data. It is interesting to note that the amplitude of the series seems to be higher in periods when the first cycle, which we associate with the business cycle, is negative. Although we do not formally test for it, it reveals that seasonality seems to be stronger in periods of contraction than in expansions.

For each draw from the posterior density of the model parameters one can use standard state space methods for simulating an out-of-sample path of values for the states and the data, consistent with the in-sample data and the parameter values. By doing so for each parameter draw from the MCMC procedure and for each model using the model weights, one effectively
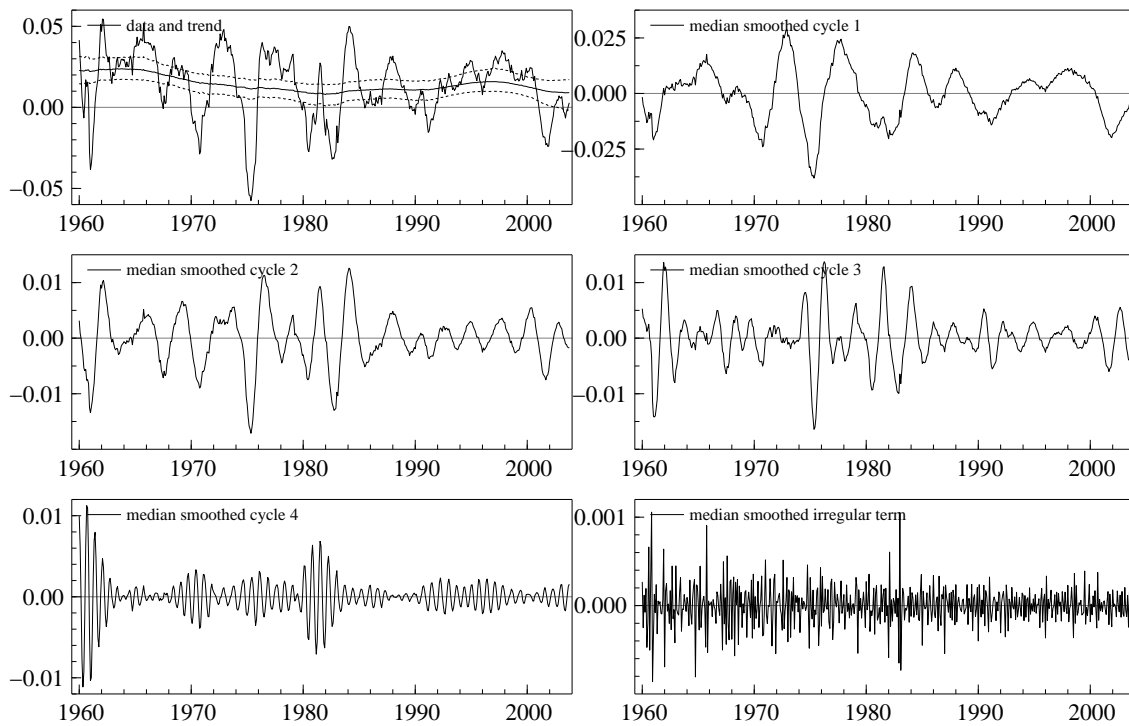
Figure 6: US Industrial Production, monthly data from 1960:1 to 2003:10. Posterior median and 5% and 95% quantiles of smoothed trend component. Also posterior medians of the other components.
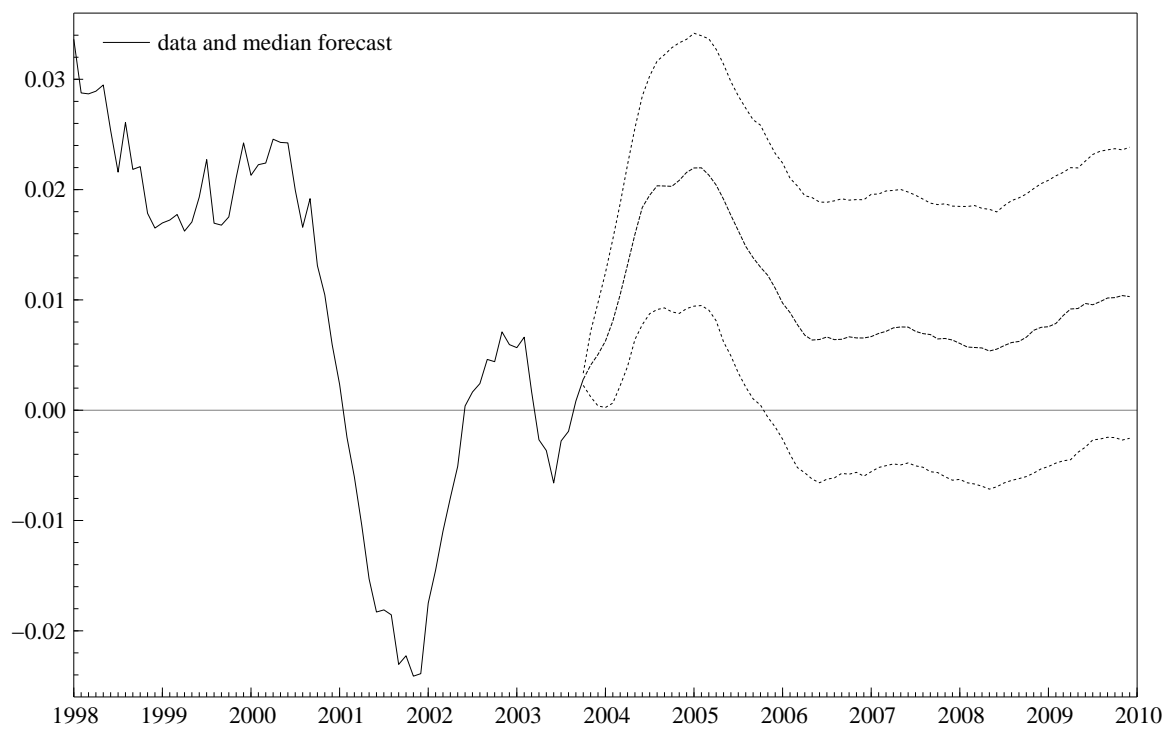
Figure 7: Forecasts of US Industrial Production. Posterior median and first and third quartiles of forecasts up to 2010.

obtains a full forecast density that incorporates all uncertainty arising from 3 sources, viz. the forecast uncertainty from the unknown future innovations, the estimation uncertainty of the model parameters and the uncertainty on the number of cycles in the model. The median and first and third quartiles of such a forecast distribution, are given in Figure 7.

4 US Unemployment

The unemployment rate is one of the most closely watched economic indicators. Its behavior is known to be asymmetric, typically exhibiting a long and slow decline followed by a short but steep increase. For policy makers it is of interest to know how much of the unemployment is due to structural causes and how much can be attributed to cyclical and seasonal factors. The variability in each is also of interest.

The issues that we will address with the decomposition in cycles are

- Can the model deal with the asymmetry?

- How does the model distinguish between permanent and transitory components of unemployment?

- Is the mean unemployment rate constant over time?

- Is the mean amplitude of the cycle constant over time?

We use monthly US unemployment rates starting 1959:01.

Table 3: Posterior model probabilities for the US Unemployment rate.

	#cycles in estimated model					
trend	1	2	3	4	5	total
no	0.000	0.006	0.064	0.067	0.264	0.401
yes	0.001	0.020	0.150	0.337	0.091	0.599

Note: Posterior probabilities under equal prior probabilities.

The estimated model probabilities are reported in Table 3. These indicate posterior odds of 60/40 in favor of a time-varying mean. The favored model contains a stochastic trend and 4 cycles. As before we calculated the implied posterior spectrum. The spectrum has 4 peaks as can be seen from Figure 8. The first 3 peaks are more clearly discernable than the fourth. The frequencies of the modes are 0.018, 0.054, 0.094, and 0.162. These correspond to periods of 350, 117, 67, and 39 months, respectively. Note that the frequency of the second cycle is
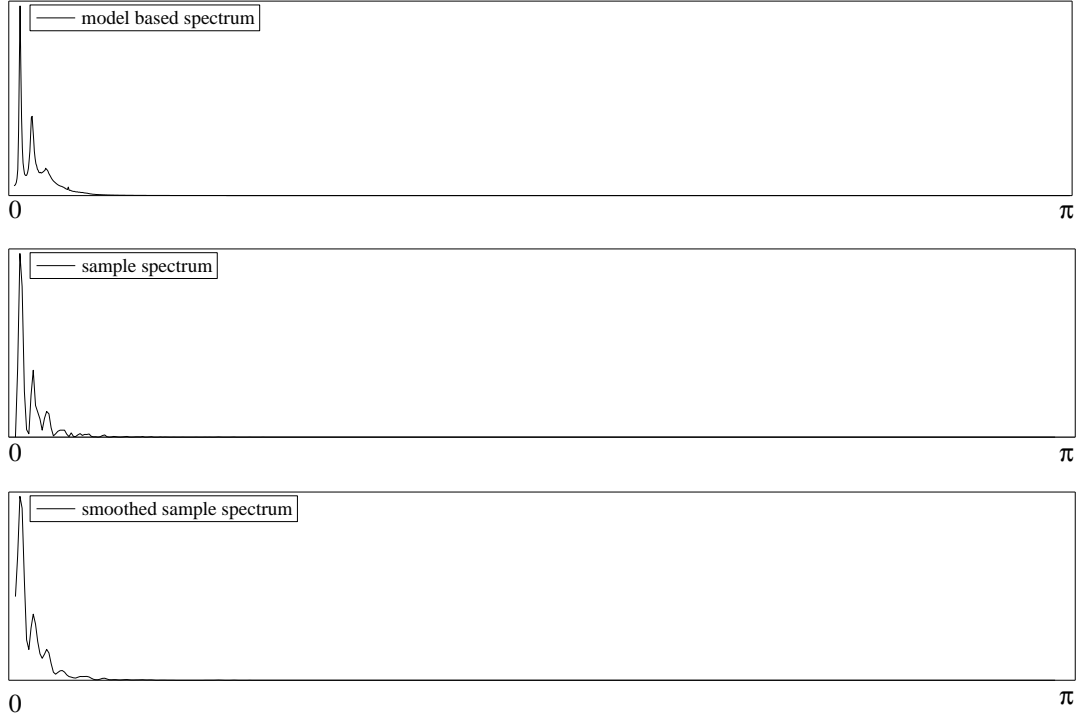
Figure 8: Sample spectra and model based spectrum of stationary components for the US unemployment rate data.

exactly 3 times the frequency of the first cycle and the fourth cycle's frequency is exactly 9 times as large. It is known that the superposition of a base cycle together with its higher order harmonics (at exact multiples of the base frequency), can generate a large variety of interesting wave forms, such as block waves, triangular waves or sawtooth waves, even when the constituents are all pure sine waves. The sawtooth wave is an extreme example of an asymmetric shape, where a single cycle consists of a gradual increase, followed by a instantaneous drop. It emerges that such a mechanism is at action in the estimated model for the unemployment data, where the asymmetry is mimiced by 3 cycles.

The smoothed states are plotted in Figure 9. The unobserved time-varying mean provided by the stochastic trend component measures the permanent part of the unemployment rate. It varies only slowly over time and has a value of 6.8% at the end of the observation period. The latent amplitude of the c-th stochastic cycle at time t is given by $\sqrt{\psi_{ct}^2 + \psi_{ct}^{*2}}$. The total latent amplitude of the cycle, which is given by the sum of the amplitudes over the cycles, is the potential deviation when all cycles are simultaneously at their maximum value (or all at their minimum). Of course, the unemployment rate cannot take on negative values. The amplitude thus serves as a measure of cyclical variability. Figure 10 shows the median and the
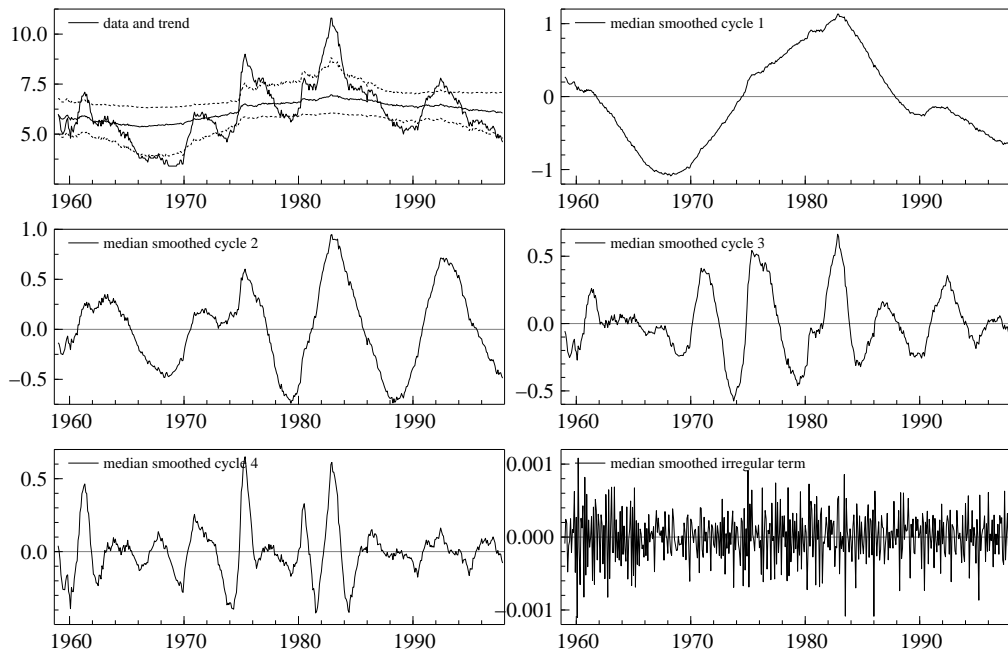
21

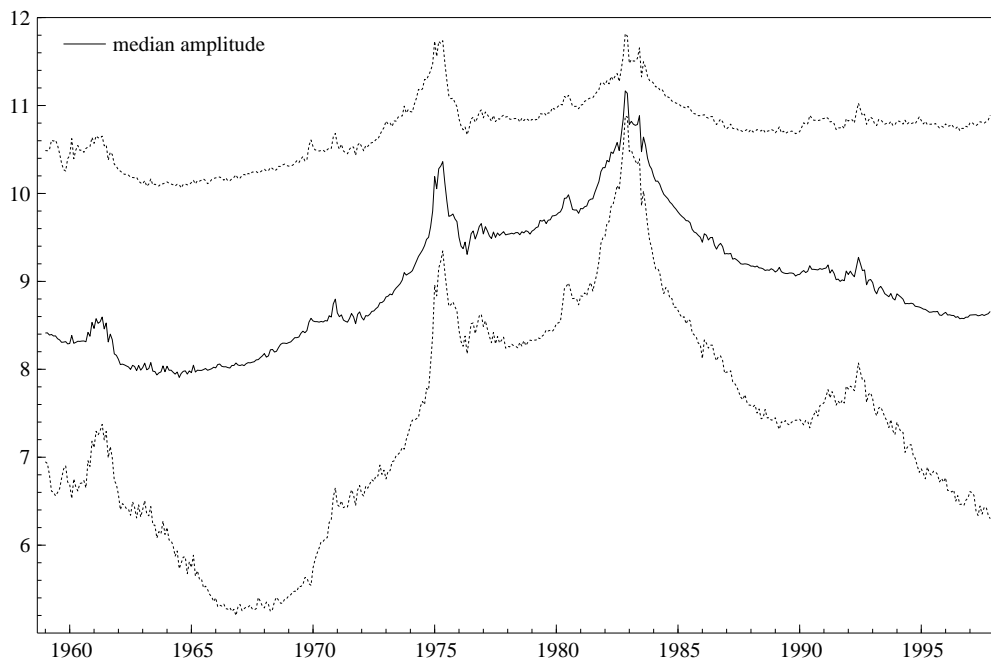Figure 9: Smoothed states for the US unemployment rate data.

Figure 10: Smoothed joint amplitude of all cycles for the US unemployment rate data.

5% and 95% quantiles of the posterior distribution of the total smoothed latent amplitudes of the cyclical components at each point in time. The striking feature in the amplitudes are the peaks that occur only when the original data series switches from increasing to decreasing and not when the unemployment rate switches from decreasing to increasing. Presumably, the number of harmonics is sufficient to capture the latter transition but it is insufficient for describing the richer harmonics needed to describe the sudden downturn of the unemployment rates. In that case the flexibility of the stochastic cycles suffices to fit the harmonic dynamics.

Another tendency that can be observed from Figure 10 is the decline in amplitude that has occurred since 1985 to the end of the observation period back to the levels of the 1960-s. We view this as an indication that the unemployment rate has become less cyclical.

5   Real Exchange Rates

Parity conditions play a crucial role in international economics. These intuitive rules provide theoretical insight in the mechanisms which govern international markets. Furthermore, these parity relations have important empirical relevance since they provide structure for modelling exchange rates, prices and interest rates in a multi-country setting. Purchasing Power Parity (PPP), for instance, links exchange rates to price levels. In its simplest version it states that the price of a good purchased locally in the domestic currency should be the same as the price of the same good purchased abroad in the foreign currency after taking the exchange rate into account. The hypothesis is usually extended to compare (changes in) price indices instead of single goods and to allow for transitory deviations from a single fixed price. The PPP hypothesis implies that the time series of the real exchange rate follows a stationary process. There exists a huge literature on testing the stationarity properties of real exchange rates. A diversity of models and techniques have been applied, including for example regression models, unit root tests, stationarity tests, cointegration, and panel methods. Earlier Bayesian work on PPP includes Schotman and van Dijk (1991) who use an autoregressive model with a fixed mean. One of the most striking findings in this area during the last decade has been the so-called Purchasing Power Parity Puzzle, see Rogoff (1996). Real exchange rates typically show huge variability in the short to medium term and at the same time it has a very high persistence with a half-life of deviations from its mean of 3 to 5 years. The high volatility could potentially be explained by sticky prices and real shocks, but the mean reversion in the data is too slow to be consistent with these arguments. Considering the major importance of PPP as a corner-stone in many international macro models, the failure to find convincing evidence in real-world data for PPP is one of the most intriguing disparities between economic theory and empirical data. It is not only of theoretical interest, but is also highly relevant as an empirical topic. In this section, we apply our methods to real exchange rate data and

assess its stationarity properties both in terms of posterior probabilities of a stochastic trend and in terms of the full posterior distribution of the half-life, which is a commonly reported persistence measure.

We study real exchange rates (in logs) of the German Mark with the US Dollar as numeraire currency (DEM/USD). The data on real exchange rates are obtained from the OECD Statistical Compendium. The series are constructed from nominal exchange rates and consumer price indices and have a monthly frequency covering the post-Bretton Woods period, from 1973:01 to 1998:12, when the internal Euro rates became fixed.

Table 4: Posterior model probabilities for the DM/USD real exchange rate.

	#cycles in estimated model					
trend	1	2	3	4	5	total
no	0.000	0.000	0.036	0.233	0.191	0.460
yes	0.000	0.000	0.062	0.463	0.014	0.540

Note: Posterior probabilities under equal prior probabilities.

Table 4 reports the model probabilities. Approximately equal posterior weights are assigned to the models with and without stochastic trend. This is consistent with earlier findings that it is hard to find evidence for PPP.

The estimated spectra are shown in Figure 11. The modes of the periods of the cycles are 186, 97, 52 and 11 months, respectively, for the DEM/USD real exchange rate. The period of the shorter cycle may be related to a seasonal or calender effect. The 2 intermediate frequencies may have a connection to the business cycle. The longer cycle accommodates to some medium-term dynamics. However, including the cycle components serves merely as a means of providing the model with the required flexibility to capture any relevant transitory dynamics. Attaching a structural or economic interpretation to the individual cycles is again not our primary focus.

Engel (2000) suggests that instead of temporary deviations from a fixed long-term mean it is more plausible to assume that the equilibrium rate itself is slowly changing over time. He explains this by distinguishing between traded goods for which price adjustment is likely and non-traded goods which lack an identifiable mechanism for mean reversion. Our model can be used to distinguish between permanent and transitory components. The stochastic trend plays the role of time-varying mean and the cycles, which by definition are 0 on average, form the transitory components. Note that the trend components is only a statistical construct with the desired permanency properties and should not be regarded as an estimate of the price difference of non-traded goods among countries.

Figure 11: Sample spectra and model based spectrum of stationary components for the Real exchange rate data.

In the PPP literature there is a growing interest in the half-life of deviations from PPP. The half-life is defined as the number of years before the effect of a shock becomes permanently less than half the size of the original shock and therefore has a natural connection with impulse response functions. It is an important summary statistic of the full impulse response function. The half-life is often found to be between three and five years, see e.g. Cheung and Lai (2000) and the references cited therein. Note that this finding is based on models that impose PPP in a rigid way. Cheung and Lai (2000) conclude that classical point estimates of the half-life have a very large imprecision. They construct confidence intervals to quantify the uncertainty. In this paper we also find non-monotonic impulse responses, consistent with their findings.

In our model a fixed mean is not imposed. In our model, we first calculate the impulse response function of the deviations from the (time-varying) mean by means of the infinite moving average or Wold representation, see e.g. Hamilton (1994),

$$\Psi_i = ZT^{i-1}K, \tag{5.1}$$

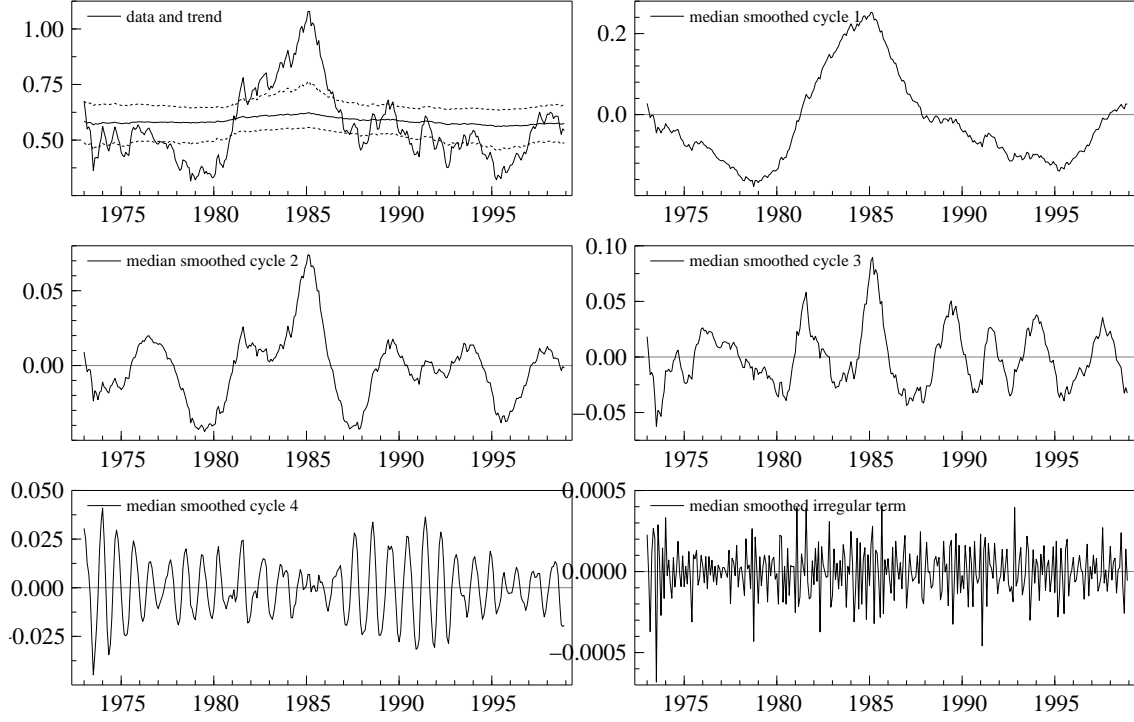where Z is the selection vector in the measurement equation, T is the block diagonal state

Figure 12: DM/USD Real exchange rate, monthly data from 1973:1 and posterior median of smoothed states.

transition matrix with blocks

$$\rho_c \begin{pmatrix} \cos \omega_c \pi & \sin \omega_c \pi \\ -\sin \omega_c \pi & \cos \omega_c \pi \end{pmatrix}$$

and K is the steady-state Kalman gain which is straightforwardly obtained from the Kalman filter. A unit shock is then assumed for $t = 0$. It is important to note that it matters which error term in our model is subjected to a shock. The value of the frequency parameter and the damping factor of that cycle influence how the effect of the shock will evolve over time. An unambiguous way of attributing the initial shock to the different components is to do so according to the Kalman gain. This provides an intuitive explanation of (5.1). Given the impulse response function, the half-life is trivially the time it takes for the impulse response to fall permanently below a half. Different possible parameter configurations result in a variety of impulse responses. The Bayesian methodology enables us to integrate out the parameter uncertainty, weighting each possible impulse response and its implied half-life according to the posterior probability of the particular outcome. Note that the posterior expectation of the half-life is not the same as the half-life of the expected impulse response, because the half-life is a highly nonlinear function of the impulse response. In Figure 13 we see that the posterior
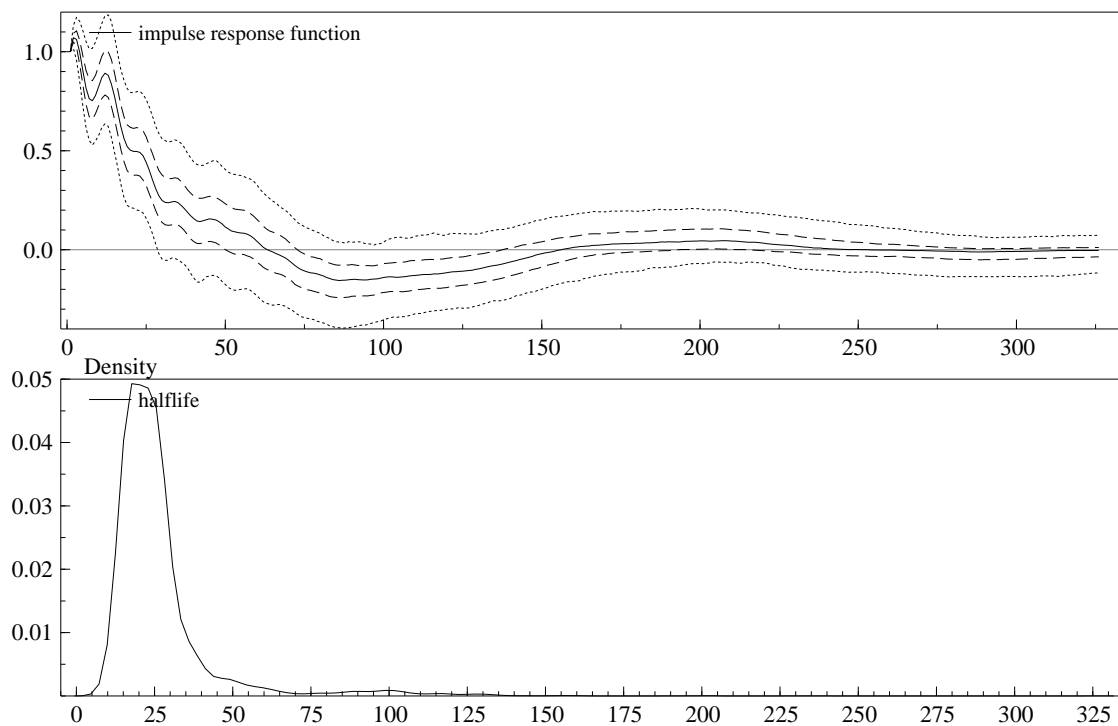
Figure 13: Posterior median and 5%, 25%, 75% and 95% quantiles of the impulse response function and density of the halflife for the DM/USD real exchange rate data.

density of the half-life even almost entirely excludes half-lives of 3 to 5 years. The half-lives we find with a median value of 23 months are more likely to be attributable to sticky prices and other rigidities than the earlier findings. Including a flexible mean is the main reason that we find half-lives under two years for both real exchange rates. This is much shorter than the half-lives reported in previous studies which are in the range of 3 to 5 years. Note that we cannot use the full model to calculate half-lives. In that case the impulse responses are related to shocks to the entire system instead of shocks to the deviations from the mean. Part of the shock will be attributed to the trend component with the effect that the slope is changed. Thus, the initial shock will have a permanent and even increasing effect over time. The impulse response function will therefore diverge and no finite half-life exists in that case.

6 Conclusion

We have introduced a class of models, which consist of a stochastic trend, an *unknown* number of non-deterministic cyclical components and an irregular term. This specification leads to a very flexible functional form, which can capture a wide range of dynamics in the data. A

MCMC-based Bayesian procedure is provided for sampling from the posterior distribution of the model parameters. Based on these samples, the posterior model probabilities can be calculated, indicating the appropriate number of cycles in the model. The parameter uncertainty and the uncertainty on the number of cycles can be fully taken into account in the subsequent inference. Prior specification is facilitated by a reparametrization and an efficient candidate generator for the model parameters of unobserved cycle components accelerates the MCMC-scheme.

Simulated data were used to illustrate model choice, model averaging, and that noise fitting is avoided. We have provided estimates of the cycles in US Industrial Production, extracted the evolution of the average growth rate over time and calculated forecast densities that incorporate all uncertainty originating from several sources. The asymmetry of the cycles in the US Unemployment Rate data are adequately described by cyclical components at some selected higher order harmonic frequencies. Also, the mean unemployment rate and the amplitude of the cyclical variability were calculated. For the DM/USD real exchange rate, we investigated the stationarity properties. Further, we reported the impulse response function and the posterior density of the half-lives of deviation from the time-varying mean. The resulting half-lives are more plausible than the values commonly reported in the literature.

A   Posterior sampler for a single cycle component

Consider one of the cycle components, i.e.:

$$\begin{pmatrix} \psi_t \\ \psi_t^* \end{pmatrix} = \rho \begin{pmatrix} \cos\omega\pi & \sin\omega\pi \\ -\sin\omega\pi & \cos\omega\pi \end{pmatrix} \begin{pmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{pmatrix} + \begin{pmatrix} \kappa_t \\ \kappa_t^* \end{pmatrix}$$

$$= \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \begin{pmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{pmatrix} + \begin{pmatrix} \kappa_t \\ \kappa_t^* \end{pmatrix},$$

with $a = \rho\cos\omega\pi$ and $b = \rho\sin\omega\pi$ and inversely $\rho = \sqrt{a^2 + b^2}$ and $\omega = \frac{1}{\pi}\arctan\frac{b}{a}$.

In matrix notation and takings the vec at both sides

$$\begin{pmatrix} \Psi & \Psi^* \end{pmatrix} = \begin{pmatrix} \Psi_{-1} & \Psi_{-1}^* \end{pmatrix} \begin{pmatrix} a & -b \\ b & a \end{pmatrix} + \begin{pmatrix} \kappa & \kappa^* \end{pmatrix}$$

$$\begin{pmatrix} \Psi \\ \Psi^* \end{pmatrix} = \left(I_2 \otimes \begin{pmatrix} \Psi_{-1} & \Psi_{-1}^* \end{pmatrix} \right) vec \begin{pmatrix} a & -b \\ b & a \end{pmatrix} + \begin{pmatrix} \kappa \\ \kappa^* \end{pmatrix}$$

$$= \left(I_2 \otimes \begin{pmatrix} \Psi_{-1} & \Psi_{-1}^* \end{pmatrix} \right) \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} \kappa \\ \kappa^* \end{pmatrix}$$

$$= \begin{pmatrix} \Psi_{-1} & \Psi_{-1}^* \\ \Psi_{-1}^* & -\Psi_{-1} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} \kappa \\ \kappa^* \end{pmatrix}$$

This takes algebraically the form of a simple linear regression

$$y = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma_\kappa^2 I_{2T})$$

with the OLS estimator given by $\hat{\beta} = (X'X)^{-1}X'y$ given by

$$\frac{1}{\Psi_{-1}'\Psi_{-1} + \Psi_{-1}^{*}{'}\Psi_{-1}^*} \begin{pmatrix} \Psi_{-1}'\Psi + \Psi_{-1}^{*}{'}\Psi^* \\ \Psi_{-1}^{*}{'}\Psi - \Psi_{-1}'\Psi^* \end{pmatrix}$$

The Jacobian of the transformation:

$$\begin{pmatrix} a \\ b \\ \sigma_\kappa^2 \end{pmatrix} = \begin{pmatrix} \rho\cos\omega\pi \\ \rho\sin\omega\pi \\ (1-\rho^2)\sigma_\psi^2 \end{pmatrix}$$

is

$$\left| \frac{\partial \begin{pmatrix} a & b & \sigma_\kappa^2 \end{pmatrix}'}{\partial \begin{pmatrix} \rho & \omega & \sigma_\psi^2 \end{pmatrix}} \right| = \begin{vmatrix} \cos\omega\pi & -\rho\pi\sin\omega\pi & 0 \\ \sin\omega\pi & \rho\pi\cos\omega\pi & 0 \\ -2\rho\sigma_\psi^2 & 0 & 1-\rho^2 \end{vmatrix} = (1-\rho^2)\rho\pi$$

B   Technical Details of the Posterior Simulator

- We primarily use the parametrization characterized by the parameters $\theta = (\{\sigma_i^2, \omega_i, \rho_i\}, \sigma_\xi^2, \sigma_\eta^2)$

- For the prior specification we use the hierarchical parametrization. When the prior densities are needed in the posterior sampler, we change to the model parametrization by the transformation $(\{\lambda_i\}, \sigma^2, \nu)$ to $(\{\sigma_i^2\}, \sigma_\xi^2, \sigma_\eta^2)$

- The candidates in the Metropolis-Hastings step are drawn from the natural conjugate linear model in the candidate parametrization, and converted from $\{a_i, b_i, \sigma_{\kappa_i}^2\}$ to $\{\sigma_i^2, \omega_i, \rho_i\}$.

The Gibbs sampler consists of the following steps

1. $p(\{\Psi_i, \Psi_i^*\}, \mu | \sigma^2, \omega, \rho, \sigma_\xi^2, \sigma_\eta^2, y)$ by a simulation smoother

2. $p(\{\sigma_i^2, \omega_i, \rho_i\}, \sigma_\xi^2, \sigma_\eta^2 | \{\Psi_i, \Psi_i^*\}, \mu, y)$using a Metropolis-Hastings step (within Gibbs):

    (a) candidate $(\sigma_i^2, \omega_i, \rho_i)$ from $q(a_i, b_i, \sigma_{\kappa_i}^2 | \Psi_i, \Psi_i^*)$ for $i = 1, \ldots, C$

    (b) candidate σ_ξ^2 from $q(\sigma_\xi^2 | \xi)$, with $\xi_t = y_t - \sum \psi_{it} - \mu_t$

    (c) candidate σ_η^2 from $q(\sigma_\eta^2 | \eta)$, with $\eta_t = \Delta \mu_t$

    (d) acceptance or rejection of all these candidates jointly.

ad 2(a-c). For the candidate (pseudo) prior the researcher faces a choice between a simple prior with a straightforward candidate sampler on one hand and a more involved prior that captures more of the structure of the target density on the other hand. This latter choice also leads to a more difficult candidate sampler. In our experience, the sampler runs fine without this extra effort. The simplest choice would be the non-informative prior taking the initial values of the state as given. In its essence this choice is not more involved than a linear regression model. Note that even the ordering of the frequencies is not specified here.

If $q(a_i, b_i, \sigma_{\kappa_i}^2) \propto (\sigma_{\kappa_i}^2)^{-1}$, then

$$q(\sigma_i^2, \omega_i, \rho_i) = q(a_i, b_i, \sigma_{\kappa_i}^2)|J_i|$$
$$= \frac{1}{(1 - \rho_i^2)\sigma_i^2}(1 - \rho_i^2)\rho_i \pi = \frac{\rho_i \pi}{\sigma_i^2}$$

For the irregular we assume $q(\sigma_\xi^2) \propto (\sigma_\xi^2)^{-1}$ and hence the total proposal prior becomes:

$$q(\theta | \alpha_0) = q(\sigma_\xi^2)q(\sigma_\mu^2)\prod q(\sigma_i^2, \omega_i, \rho_i | \Psi_{i0}, \Psi_{i0}^*)$$
$$= \frac{\pi}{\sigma_\xi^2}\prod \frac{\rho_i}{\sigma_i^2}$$

Under these candidate priors, the candidates are drawn from:

$$\frac{\xi'\xi}{\sigma_\xi^2}|\xi \sim \chi_T^2$$

$$\frac{y'M_Xy}{\sigma_{\kappa_i}^2}|\Psi_i, \Psi_i^* \sim \chi_{T-2}^2$$

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix}|\sigma_{\kappa_i}^2, \Psi_i, \Psi_i^* \sim N((X'X)^{-1}X'y, \sigma_{\kappa_i}^2(X'X)^{-1})$$

and mapped to ρ_i, ω_i, σ_i^2.

ad 2(d). The acceptance probability is given by

$$\min\left\{\frac{\left(\frac{p(\theta^*|\alpha,y)}{q(\theta^*|\alpha,y)}\right)}{\left(\frac{p(\theta^{(i-1)}|\alpha,y)}{q(\theta^{(i-1)}|\alpha,y)}\right)}, 1\right\}$$

with

$$\frac{p(\theta|\alpha,y)}{q(\theta|\alpha,y)} = \frac{L(\alpha,y|\theta,\alpha_0)L(\alpha_0|\theta)p(\theta)}{L_c(\alpha,y|\theta,\alpha_0)q(\theta|\alpha_0)}$$
$$= \frac{L(\alpha_0|\theta)p(\theta)}{q(\theta|\alpha_0)}$$

Note that the factors related to the likelihoods cancel in matching posteriors $p(\cdot)$ and $q(\cdot)$, leaving only the ratio of the prior densities and the initial condition in as determinants of the acceptance probability.

The initial conditions $\Psi_{i0}, \Psi_{i0}^*|\sigma_i^2, \omega_i, \rho_i \sim N(0, \sigma_i^2)$ lead to

$$L(\alpha_0|\theta) = L(\Psi_{\cdot0}, \Psi_{\cdot0}^*|\{\sigma_i^2, \omega_i, \rho_i\}, \sigma_\xi^2, \sigma_\mu^2) \propto \prod(\sigma_i^2)^{-1}\exp\left(-\frac{1}{2\sigma_i^2}\sum(\Psi_{i0}^2 + \Psi_{i0}^{*2})\right)$$

If the prior of the model parameters is given by

$$p(\{\lambda_i, \omega_i, \rho_i\}, \sigma^2, \nu) \propto \sigma^{-2}\nu^{-\frac{1}{2}}(1-\nu)^{-\frac{1}{2}}\prod\rho_i^2$$

then

$$p(\theta) = p(\{\sigma_i^2, \omega_i, \rho_i\}, \sigma_\xi^2, \sigma_\mu^2)$$
$$= p(\{\lambda_i, \omega_i, \rho_i\}, \sigma^2, \nu)\left|\frac{\partial\left(\sigma_1^2 \quad \cdots \quad \sigma_C^2 \quad \sigma_\xi^2 \quad \sigma_\mu^2\right)'}{\partial\left(\lambda_1 \quad \cdots \quad \lambda_C \quad \sigma^2 \quad \nu\right)}\right|^{-1}$$
$$= (\sigma^2)^{-C-2}\nu^{-\frac{1}{2}}(1-\nu)^{-C-\frac{1}{2}}\prod\rho_i^2$$

with

$$\sigma^2 = \sigma_\xi^2 + \sigma_\mu^2 + \sum \sigma_i^2$$

$$\nu = \frac{\sigma_\mu^2}{\sigma^2}$$

$$\lambda_i = \frac{\sigma_i^2}{(1-\nu)\sigma^2}$$

Here we have used that the transformation between $\begin{pmatrix} \lambda_1 \cdots \lambda_C & \sigma^2 & \nu \end{pmatrix}'$ and $\begin{pmatrix} \sigma_1^2 \cdots \sigma_C^2 & \sigma_\xi^2 & \sigma_\mu^2 \end{pmatrix}'$ is given by

$$
\begin{pmatrix} \sigma_1^2 \\ \vdots \\ \sigma_C^2 \\ \sigma_\xi^2 \\ \sigma_\mu^2 \end{pmatrix}
=
\begin{pmatrix} (1-\nu)\lambda_1\sigma^2 \\ \vdots \\ (1-\nu)\lambda_C\sigma^2 \\ (1-\nu)(1-\sum\lambda_i)\sigma^2 \\ \nu\sigma^2 \end{pmatrix},
$$

and it has Jacobian

$$
\left| \frac{\partial \begin{pmatrix} \sigma_1^2 & \cdots & \sigma_C^2 & \sigma_\xi^2 & \sigma_\mu^2 \end{pmatrix}'}{\partial \begin{pmatrix} \lambda_1 & \cdots & \lambda_C & \sigma^2 & \nu \end{pmatrix}} \right|
=
\begin{vmatrix}
(1-\nu)\sigma^2 & & 0 & (1-\nu)\lambda_1 & -\lambda_1\sigma^2 \\
& \ddots & & \vdots & \vdots \\
0 & & (1-\nu)\sigma^2 & (1-\nu)\lambda_C & -\lambda_C\sigma^2 \\
-(1-\nu)\sigma^2 & \cdots & -(1-\nu)\sigma^2 & (1-\nu)(1-\sum\lambda_i) & -(1-\sum\lambda_i)\sigma^2 \\
0 & \cdots & 0 & \nu & \sigma^2
\end{vmatrix}
$$

$$
= \begin{vmatrix} P & R \\ S & Q \end{vmatrix} = |Q||P - RQ^{-1}S| = |P||Q - SP^{-1}R| = |P|\sigma^2
$$

$$
= (1-\nu)^C(\sigma^2)^{C+1}.
$$

References

C.K. Carter and R. Kohn. On Gibbs sampling for state space models. *Biometrika*, **81**: 541–553, 1994.

G. Casella and E.I. George. Explaining the Gibbs sampler. *The American Statistician*, **46**: 167–174, 1992.

Y-W. Cheung and S.L. Lai. On the purchasing power parity puzzle. *Journal of International Economics*, **52**:321–330, 2000.

S. Chib. Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, **90**(432):1313–1321, 1995.

P. de Jong and N. Shephard. The simulation smoother for time series models. *Biometrika*, **82**:339–350, 1995.

C. Engel. Long-run PPP may not hold after all. *Journal of International Economics*, **57**: 243–273, 2000.

S. Frühwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, **15**:183–202, 1994.

A.E. Gelfand and A.F.M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:389–409, 1990.

J. Geweke. Using simulation methods for Bayesian econometric models: inference, development and communication. *Econometric Reviews*, **18**(1):1–73, 1999.

J.D. Hamilton. *Time Series Analysis*. Princeton University Press, Princeton, NJ, 1994.

A.C. Harvey. *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, Cambridge, 1989.

A.C. Harvey. *Time series models*. Harvester Wheatsheaf, 2nd edition, 1993.

A.C. Harvey and M. Streibel. Testing for deterministic versus indeterministic cycles. *Journal of Time Series Analysis*, **19**:505–529, 1998.

A.C. Harvey, T.M. Trimbur, and H.K. Van Dijk. Cyclical components in economic time series: a Bayesian approach. Econometric Institute Report EI2002-20, Erasmus University Rotterdam, 2002. available at http://www.econometric-institute.org.

G. Huerta and M. West. Priors and component structures in autoregressive time series models. *Journal of the Royal Statistical Society* **B**, **61**:881–899, 1999.

G. Koop and H.K. van Dijk. Testing for integration using evolving trend and seasonals models: A Bayesian Approach. *Journal of Econometrics*, **97**:261–291, 2000.

K. Rogoff. The purchasing power parity puzzle. *Journal of Economic Literature*, **34**:647–668, 1996.

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

M.A. Tanner and W.H. Wong. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, **82**:528–550, 1987.

M. West. Bayesian inference in cyclical component dynamic linear models. *Journal of the American Statistical Association*, **90**:1301–1312, 1995.

E. Zivot. A Bayesian analysis of the unit root hypothesis within a unobserved components model. *Econometric Theory*, **10**:552–578, 1994.