Changing Correlation and Portfolio Diversification Failure in the Presence of Large Market Losses

Alessio Sancetta and Steve E. Satchell

March 2003

CWPE 0319

Cyclical components in economic time series: a Bayesian approach

Andrew C. Harvey[a], Thomas M. Trimbur[a], Herman K. van Dijk[b]
[a] Cambridge University, Faculty of Economics and Politics
[b] Econometric Institute, Erasmus University

October 28, 2002

Abstract

Cyclical components in economic time series are analysed in a Bayesian framework, thereby allowing prior notions about periodicity to be used. The method is based on a general class of unobserved component models that allow relatively smooth cycles to be extracted. Posterior densities of parameters and smoothed cycles are obtained using Markov chain Monte Carlo methods. An application to estimating business cycles in macroeconomic series illustrates the viability of the procedure for both univariate and bivariate models.

KEYWORDS: Band pass filter, Gibbs sampler, Kalman filter, Markov chain Monte Carlo, state space, unobserved components.

JEL classification: C11, C32, E32

1 Introduction

Decomposing time series into trends and cycles is fundamental to a good deal of macroeconomic analysis. The Hodrick-Prescott (HP) filter is often used to detrend series but as shown in Harvey and Jaeger (1993) and Cogley and Nason (1995), using it inappropriately can result in the creation of spurious cycles. The same is true of the band pass filter recently proposed by Baxter and King (1999) for extracting cyclical movements over the range two to ten years; see Murray (2002).

Harvey and Jaeger (1993) argued that detrending is best accomplished by fitting a structural time series model consisting of trend, cycle and irregular unobserved components. The model is estimated in state space form with the components extracted by the Kalman filter and associated smoother. However, fitting the model to series like GDP usually results in the irregular component disappearing with the result that the cycle is quite noisy. To overcome this feature, Harvey and Trimbur (2002) introduced an extended class of cyclical components. The higher order cycles in this class yield implicit filters that concentrate on extracting relatively more power from a narrower band of frequencies. More high frequency noise is forced into the irregular thereby yielding a smoother cycle. An ideal band pass filter emerges as a limiting case.

The parameters in structural time series models are usually estimated by maximum likelihood (ML) using the Kalman filter. Harvey and Trimbur (2002) found that this works well for series like investment where the cycle is pronounced. However, difficulties were experienced with GDP, in that the likelihood surface often appeared to exhibit irregularities thereby resulting in implausible parameter estimates. If the period of the cycle was fixed, an acceptable cycle could be extracted, but making such a strong restriction is not altogether desirable. This experience provides an incentive for exploring a Bayesian approach. Since prior information on the periods likely to characterise business cycles is readily available, it may be used to construct a prior distribution for the frequency parameter in the cyclical component. Prior information on other parameters may also be used, though this will typically be rather more vague.

The Bayesian approach is implemented by adapting a Markov chain Monte Carlo (MCMC) algorithm for the computation of posterior and predictive distributions. This leads to the complete distribution of the cyclical component and allows the computation of marginal posteriors of the trend and cyclical component at different points in time. Confidence bands can be constructed around cyclical estimates and a forecast distribution for the cyclical component presented.

Bayesian treatment of unobserved components time series models has not, to the best of our knowledge, been concerned with cycles. Fruewirth-Schnatter (1994) and Koop and van Dijk (2000) analyse trends and seasonals in various macroeconomic series, while Bos, Mahieu, and van Dijk (2000) describe a Bayesian application of time-varying volatility and heavy tails using daily financial data. The treatment of cycles leads to a number of new

technical issues which we address in this paper.

The paper is arranged as follows. Section 2 describes the extension of the class of structural time series models to include higher order cyclical components. The Bayesian treatment is developed in section 3, while section 4 illustrates the methods with applications to some US macroeconomic series. The methods and application are extended to bivariate models in section 5. Section 6 concludes. Technical details on the state space form and the Markov chain Monte Carlo algorithm are laid out in the appendix.

2 Cyclical and trend components in time series

We consider a class of unobserved component (UC) models in which the observations, $y_t, t = 1, ..., T$ are made up of a nonstationary trend component $\mu_{m,t}$, a cyclical component $\psi_{n,t}$, and an irregular term ε_t. Thus:

$$y_t = \mu_{m,t} + \psi_{n,t} + \varepsilon_t, \quad t = 1, ...T, \tag{1}$$

where the irregular is white noise, that is $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$, the $m-th$ order stochastic trend is

$$\mu_{1,t} = \mu_{1,t-1} + \zeta_t, \zeta_t \sim WN(0, \sigma_\zeta^2) \tag{2}$$

$$\mu_{i,t} = \mu_{i,t-1} + \mu_{i-1,t-1}, \quad i = 2, ..., m \tag{3}$$

while the $n-th$ order cycle is defined by

$$\begin{aligned} \begin{bmatrix} \psi_{1,t} \\ \psi_{1,t}^* \end{bmatrix} &= \rho \begin{bmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{bmatrix} \begin{bmatrix} \psi_{1,t-1} \\ \psi_{1,t-1}^* \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix}, \\ \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix} &\sim WN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\kappa^2 & 0 \\ 0 & \sigma_\kappa^2 \end{bmatrix} \right) \end{aligned} \tag{4}$$

$$\begin{bmatrix} \psi_{i,t} \\ \psi_{i,t}^* \end{bmatrix} = \rho \begin{bmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{bmatrix} \begin{bmatrix} \psi_{i,t-1} \\ \psi_{i,t-1}^* \end{bmatrix} + \begin{bmatrix} \psi_{i-1,t-1} \\ \psi_{i-1,t-1}^* \end{bmatrix}, \quad i = 2, ..., n \tag{5}$$

The parameter λ_c denotes frequency in radians while ρ is a damping factor lying between zero and one; if it is equal to one, the cycle is nonstationary.

The disturbances driving the trend and cycle are assumed to be uncorrelated with each other and with the irregular. The specification in (1) is well-suited for trend-cycle decompositions, but the model can be easily augmented, for example by adding a seasonal component.

The second-order trend, $m = 2$, is an integrated random walk; see Harvey (1989), Kitagawa and Gersch (1996), and Young (1984). This is the usual choice since it produces relatively smooth trends that allow a cycle to be separated out more easily.

The nth order stochastic cycle $\psi_{n,t}$ has periodic movements centered around a frequency of λ_c. The stochastic movements stem from the two disturbances, κ_t and κ_t^* in (4). Suppressing κ_t^* in the model specification yields a class of band pass filters that generalises the Butterworth class of filters; see Gomez (2001) and Harvey and Trimbur (2002). However, our preference here is to work with the 'balanced form' of (4); for $n = 1$ this is identical to the stochastic cycle in (Harvey 1989, p. 39) and analytical expressions for key properties are available for all n. There is, for instance, a complete closed-form expression for the covariance matrix of the cyclical state; see appendix B and Trimbur (2002) for details. This is required for initialising the Kalman filter and in the Bayesian approach it plays a key role in the Gibbs sampler, as it affects the form of the conditional posteriors. Direct computation is important as it avoids the need for a large matrix inversion when n is big.

The variance of the cycle, σ_ψ^2, is equal to $\sigma_\kappa^2/(1-\rho^2)$ for $n = 1$ and

$$\sigma_\psi^2 = \frac{(1+\rho^2)}{(1-\rho^2)^3}\sigma_\kappa^2 \tag{6}$$

for $n = 2$. The autocorrelation function for $n = 2$ is

$$\rho(\tau) = \rho^\tau \cos(\lambda_c \tau)\left[1 + \frac{(1-\rho^2)}{1+\rho^2}\tau\right], \qquad \tau = 0, 1, 2, ...$$

as compared with $\rho(\tau) = \rho^\tau \cos(\lambda_c \tau)$ in the first order model.

Analysis in the frequency domain provides complementary insight about periodic behaviour. It is shown in Trimbur (2002) that the spectral density for the second order cyclical is

$$f_\psi(\lambda) = \frac{(1-\rho^2)^3 (n_0 + n_1 \cos\lambda + n_2 \cos 2\lambda)}{(1+\rho^2)(d_0 + d_1 \cos\lambda + d_2 \cos 2\lambda + d_3 \cos 3\lambda + d_4 \cos 4\lambda)} \tag{7}$$

where the coefficients of the terms in the numerator and denominator are given by:

$$\begin{aligned}
n_0 &= 1+4\rho^2+\rho^4, \quad n_1 = -4\rho(1+\rho^2)\cos\lambda_c, \quad n_2 = 2\rho^2\cos 2\lambda_c \\
d_0 &= (1+4\rho^4+\rho^8) + 16\rho^2\cos^2\lambda_c(1+\rho^2(1+\cos^2\lambda_c+\rho^2)) \\
d_1 &= -8\rho\cos\lambda_c(1+2\rho^2+2\rho^4+\rho^6+4\rho^2\cos^2\lambda_c(1+\rho^2)) \\
d_2 &= 4\rho^2((1+\rho^4)+2\cos^2\lambda_c(1+4\rho^2+\rho^4)), \\
d_3 &= -8\rho^3\cos\lambda_c(1+\rho^2), \quad d_4 = 2\rho^4
\end{aligned}$$

The corresponding result for $n = 1$ is

$$f_\psi(\lambda) = \frac{(1-\rho^2)\,(1+\rho^2-2\rho\cos\lambda_c\cos\lambda)}{1+\rho^4+4\rho^2\cos^2\lambda_c-4\rho(1+\rho^2)\cos\lambda_c\cos\lambda+2\rho^2\cos 2\lambda} \tag{8}$$

Figure 1 shows the spectral density for first and second order processes with $\lambda_c = \pi/4, \rho = 0.9$. As n increases, the spectrum becomes sharper and more concentrated in a band around λ_c. In figure 2, the second order cycle has $\rho = 0.81$, while the damping factor for the basic cycle, $n = 1$, remains fixed at 0.9. There is now a compensating effect so that the height of the peak is about the same, but the spectral shape for the second order process is such that there is more mass in the centre and less in the tails. Increasing n further, with a corresponding reduction in ρ, leads to further movements in this direction.

The statistical treatment of the model is based on the state space form. The likelihood function is computed in the prediction error decomposition form using the Kalman filter. Maximum likelihood estimation is carried out by numerical maximisation with respect to the unknown parameters, $\sigma_\zeta^2, \sigma_\kappa^2, \sigma_\varepsilon^2, \rho$ and λ_c. Once this has been done, the trend and cycle may be extracted by the smoothing algorithm; see Harvey and Trimbur (2002) for further details.

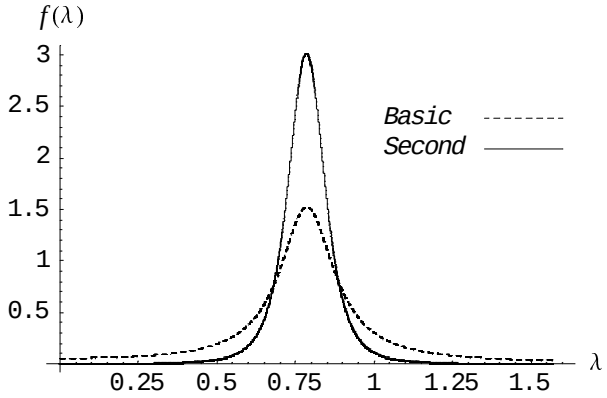

Figure 1. Spectral density of cyclical components for $n = 1$ and 2, with the parameter ρ set to 0.9.

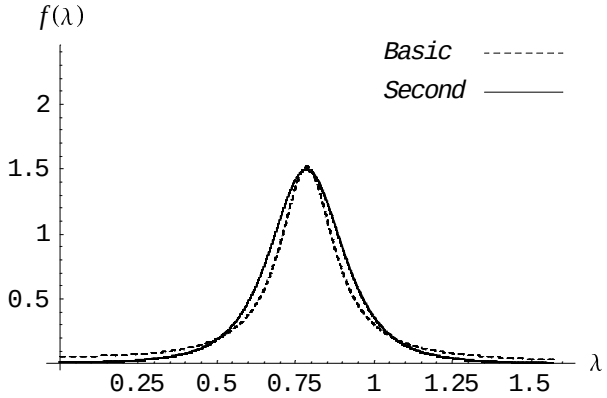

Figure 2 Spectral density of cyclical components for $n = 1$ with ρ set to 0.9 and for $n = 2$ with $\rho = 0.81$.

3 Bayesian treatment

The three variance parameters and two cyclical parameters are arranged in the vector $\theta = \{\sigma_\zeta^2, \sigma_\kappa^2, \sigma_\varepsilon^2, \rho, \lambda_c\}$. The model is assumed to have Gaussian disturbances throughout. Given a sample $Y = \{y_1, ..., y_T\}$, we specify the likelihood. Given a prior, specified below, we obtain the posterior. The goal is to analyze the properties of the posterior distribution, $p(\theta|Y)$. Since this is not a member of a class of densities which has known analytical properties, it is computed by a Markov chain Monte Carlo routine. An overview is provided here and details are given in an appendix. The method produces smoothed estimates of the cycle as a by-product.

We start by summarising the elicitation of priors. The direct interpretation of the cycle parameters, λ_c and ρ, makes it straightforward to design suitable priors; they are linked to economic intuition and previous experience of modelling business cycles. The frequency λ_c is of particular interest and the period, $2\pi/\lambda_c$, is anticipated to lie in the neighborhood of five years. A standard peaked distribution for λ_c can reflect this expectation. A practical rationale for using informative priors is that, for a moderate size sample, the likelihood surface may be unusually flat in some regions so that it becomes difficult to estimate a reasonable cycle; Harvey and Trimbur (2002) found this to be the case when modelling US GDP with a second-order cycle.

For quarterly macroeconomic data, we consider priors for λ_c centred around $2\pi/20$, based on the class of beta distributions; such priors are very flexible and are easy to work with analytically; see figure 3. The least informative prior covers a wide range of frequencies, while the sharpest density focuses attention narrowly around a period of five years. For technical details on the priors, see the appendix A.

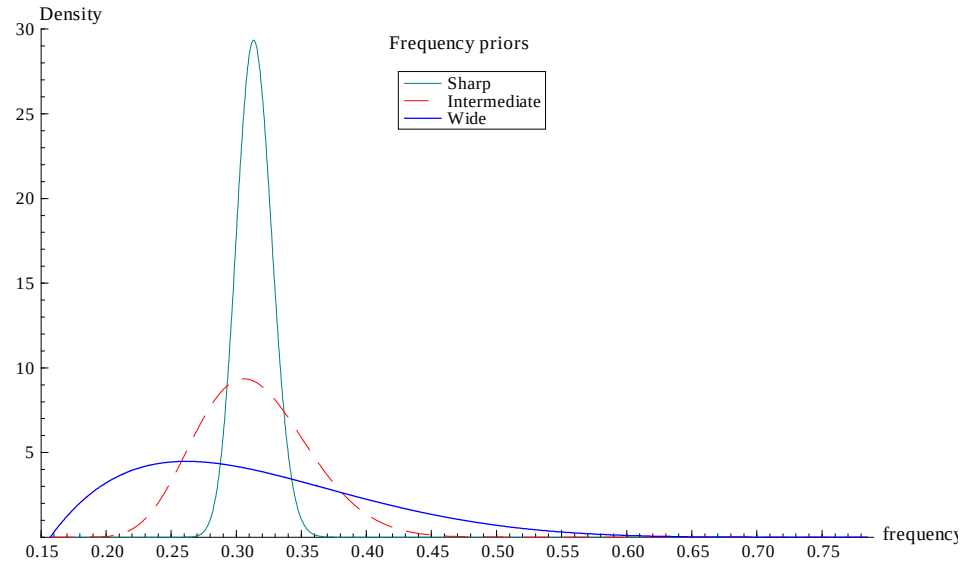

Figure 3: Beta-based priors on λ_c, with mean equal to $2\pi/20$ (five-year period for quarterly data).

The parameter ρ is linked to the order of the cycle. In the first order case ρ is the rate of decay of the cycle, but for higher orders the interpretation of ρ changes somewhat so that different values are appropriate; this is the point made by figure 2. However, since the precise form of the relationship between ρ and n is not clear, we use a uniform prior on ρ over the interval [0,1].

For all three variances, independent inverted gamma densities were used, with shape and scale set to 10^{-7} and 10^{-14}, respectively. These priors are essentially noninformative, and the use of the inverted gamma ensures proper posteriors.

Next, we summarise the evaluation of the likelihood and posterior of our unobserved components model. As normality of the disturbances is assumed, the likelihood function, $L(\theta; Y)$, can be evaluated for any permissible value of θ using the Kalman filter. This relies on the prediction error decomposition as described in Harvey (1989, p. 126). With an appropriate initialisation, the density of Y given θ is multivariate Gaussian.

The posterior $p(\theta|Y)$ is proportional to the product of the prior and likelihood. However, the expression for the product $p(\theta)L(\theta; Y)$ does not represent the kernel of a known distribution. The normalizing constant (equal to the marginal likelihood) required for evaluating the posterior ordinate is

not known in terms of elementary functions (like $\sqrt{2\pi}$ in the case of the normal distribution). A strategy is needed for analyzing the properties of $p(\theta|Y)$. With a five-dimensional parameter vector, MCMC methods offer an efficient way to sample (pseudo-random) parameter drawings from the posterior. This may be used to produce drawings of regular functions of the parameters. Thus finite sample results on posterior moments may be compared with ML estimates.

The MCMC method also produces drawings from the joint density of the two unobserved components, the cycle and trend, over the sample period. These high-dimensional variates, conditional on the data, can be used to compute a Bayesian analogue of the classical smoother. They form part of the Gibbs sampler and no additional effort is required to obtain them.

Smoothed estimates of the cycle are obtained by averaging over draws. Classical estimates correspond to the conditional mean of the cyclical component, given the sample, assuming the parameters are fixed at estimated values. The *Bayesian smoother* incorporates parameter uncertainty (the parameter vector is integrated out) and accounts for prior knowledge about the period. As for the parameters, drawings of regular functions of the components over the sample are directly obtained. This enables properties of the time-varying cycle, such as amplitude, to be studied.

Next, we briefly discuss model evaluation in our Bayesian approach. In practice, models are compared based on a particular sample y, which represents a draw from a marginal, or unconditional, density $m(Y)$ over all possible realisations Y. The predictive density $m(Y)$ shows the probabilities associated with hypothetical samples, before the data are actually observed. When evaluated at y, the marginal likelihood $m(y)$ is obtained. Higher values of $m(y)$ imply greater success in predicting the data, and therefore, the marginal likelihood offers a way to assess different models; see Kass and Raftery (1995) for a discussion. In the unobserved components framework, this allows a comparison of different cyclical orders. The marginal likelihood $m(y)$ is obtained by integrating out the parameter vector θ:

$$m(y) = \int L(\theta; y)p(\theta)d\theta \tag{9}$$

Bayes factors are computed as the ratio of marginal likelihoods. If there is an *a priori* preference for certain shapes of $p(\theta)$ or for particular component structures, this can be reflected in the prior probability ratios over the various models $\{L(\theta; y),\ p(\theta)\}$. The relative posterior probabilities associ-

ated with different model pairs can then be determined by multiplying the Bayes factors by prior probability ratios. In the current application dealing with business cycle components, there may be a preference for sharper priors on the frequency. In practice, this would mean that, unless the marginal likelihood decreases substantially when moving to a sharp density, such priors may be preferred. Some ways to estimate the marginal likelihood are discussed in the appendix.

In a Bayesian framework, forecasts account for the uncertainty in the hyperparameters. The prediction at a horizon of h quarters is

$$E(y_{t+h}|Y_T) = \int E(y_{t+h}|Y_T, \theta)p(\theta|Y)d\theta \tag{10}$$

This is the sum of the forecasts of the cyclical and trend components. The classical predictor, $E(y_{t+h}|Y_T, \theta)$, is constructed assuming θ to be fixed.

The informational content of the final posterior sample depends on the correlation between adjacent draws. By burning a large number of initial iterations and then skipping a certain number of iterations in between draws, the degree of correlation was reduced. As a result, in each case, the correlograms for the series of parameter draws declined rapidly with lag length; see the appendix for further discussion.

4 Application to macroeconomic series

This section illustrates the approach with quarterly US real GDP and investment data from 1947:1 to 2001:4 (Source: US Dept. of Commerce, Bureau of Economic Analysis: National Accounts Data), and annual US real GDP data from 1870 to 1998 compiled from the OECD publications *Monitoring the World Economy* and *The World Economy: a Millennial Perspective*. We present posterior and predictive results for the parameters for different cyclical orders for various priors and we examine smoothed cycles and trends. By using different priors we investigate the robustness of the posterior results to the choice of prior. Reference should be made to Harvey and Trimbur (2002) for the corresponding classical results.

4.1 Hyperparameters

The marginal posterior densities for $n = 1$, with the least informative prior on λ_c, are shown in figures 4 and 5. As business cycle researchers may tend to think in terms of duration or periodicity, it is interesting to look at the implied densities for the period, $2\pi/\lambda_c$. The prior and posterior for the frequency and period appear in figure 5. The posteriors show a clear peak around a five-year period even with a relatively noninformative prior.

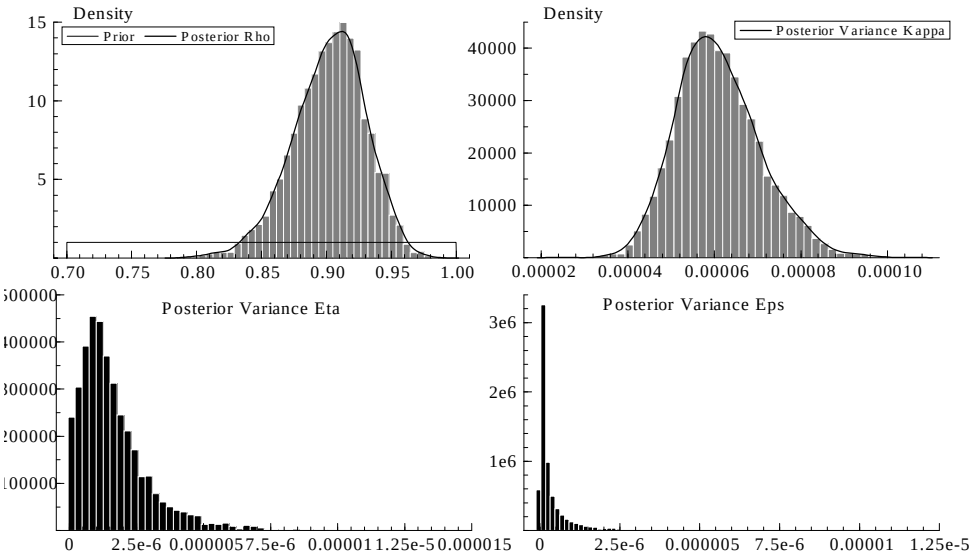

Figure 4: Marginal posterior densities of ρ, σ_κ^2, σ_ζ^2, and σ_ε^2 for $n = 1$, with least informative prior on λ_c, for quarterly US real GDP (logarithms) from 1947:1 to 2001:4.

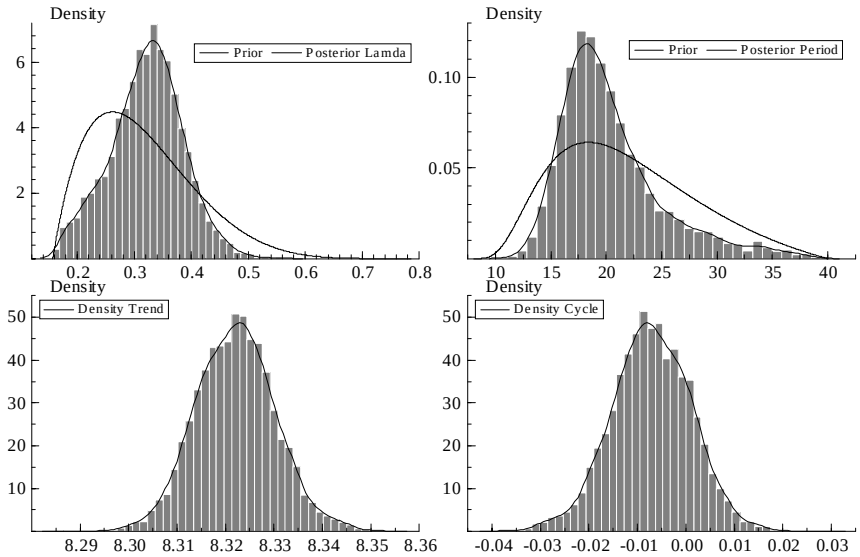

Figure 5: Marginal posterior densities of λ_c and $2\pi/\lambda_c$, and trend and cyclical components $\mu_{t,T}$ and $\psi_{t,T}$ in middle of sample ($t = 110$), for $n = 1$ with least informative prior on λ_c.

The other cyclical parameters similarly have well-behaved marginal posteriors. Based on a uniform prior on [0,1], the marginal posterior for ρ peaks near 0.9, and the density of the cyclical error variance σ_κ^2 appears symmetric. These results suggest that the likelihood surface in the first order case has a more or less regular shape so that it is straightforward to pick out a business cycle component. Note that each displayed density represents a standard approximation based on Gaussian kernels; see (Doornik 1999, p. 216). Figure 4 shows that, for the slope variance σ_ζ^2, the marginal posterior is skewed. There is significant mass near zero (the local Gaussian approximation is not shown in this case since it would, due to the skewness, extend below zero and so would not represent a meaningful density), but as shown in the histogram, the inverted gamma prior ensures that only positive values are considered.

4.2 Trends and cycles

The Bayesian analysis produces draws from the joint posterior of the trend and cyclical components, $\mu_{t,T}$ and $\psi_{t,T}$, over the sample period. The marginal posteriors of the trend and cyclical components in the middle of the sample are shown in the lower panels of figure 5. We note that the posterior mean

is the optimal estimator for a quadratic loss. The estimated trend and cycle over the entire sample period are displayed in figures 6 and 7. The shaded regions in figure 7 denote recessions as identified by the NBER. These series are obtained by averaging over the J state draws, that is

$$\widehat{\mu}_{t,T} = \frac{1}{J}\sum_{j=1}^{J}\mu_t^{(j)}, \qquad \widehat{\psi}_{t,T} = \frac{1}{J}\sum_{j=1}^{J}\psi_t^{(j)}, \quad t = 1, ..., T,$$

where $\mu_t^{(j)}$ denotes the jth draw for the trend at time t and similarly for the cyclical component. The standard deviation is estimated by $\sqrt{\sum \mu_t^{2(j)}/J - \widehat{\mu}_{t,T}^2}$ and other higher-order moments may be computed in a similar way.

The amplitude of the cycle at time t is estimated by

$$A_t = \frac{1}{J}\sum_{j=1}^{J}\sqrt{\psi_t^{2(j)} + \psi_t^{*2(j)}}, \quad t = 1, ..., T$$

The evolving amplitude is shown in figure 8. Significant time variation is evident, with especially high levels reached in the early 1950s and around the times of the oil price shocks in the 1970s.

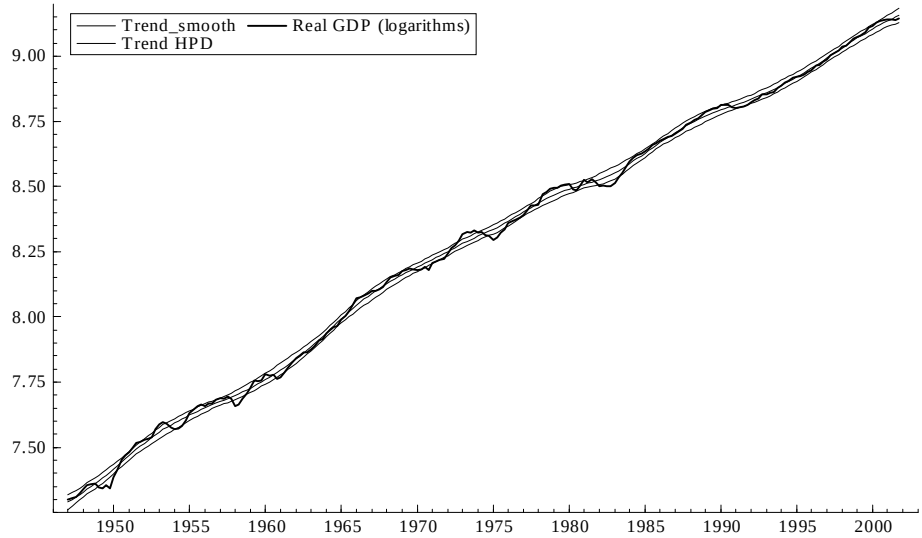

Figure 6: Estimated trend in quarterly US real GDP (logarithms) and 95% confidence bands for $n = 1$ with least informative prior on λ_c.

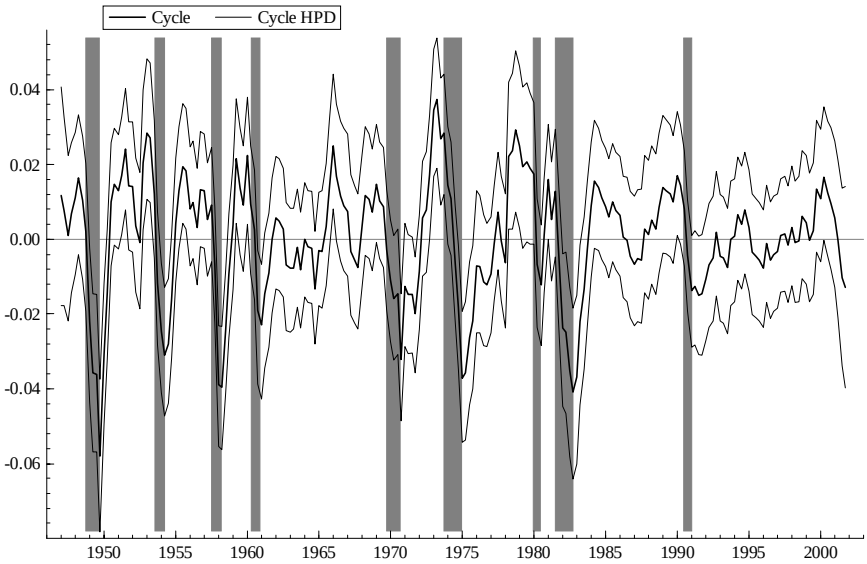

Figure 7: Estimated cycle in quarterly US real GDP (logarithms) and 95% confidence bands for $n = 1$ with least informative prior on λ_c.

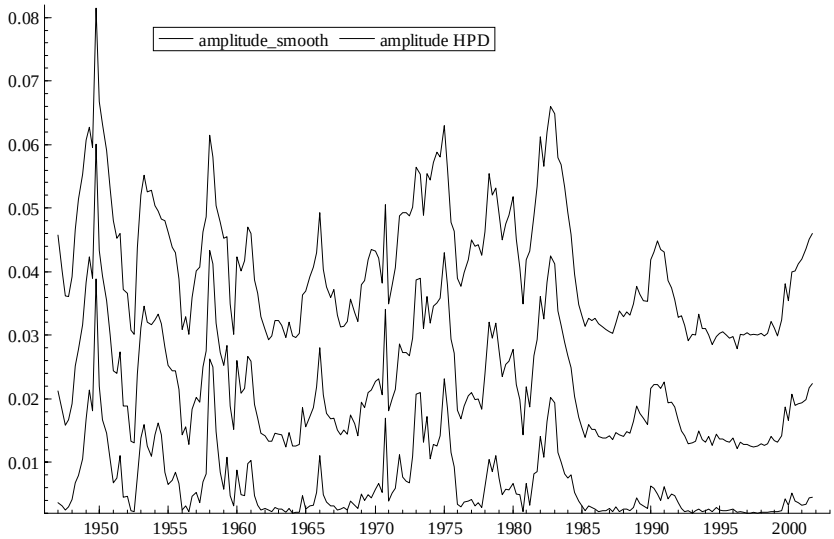

Figure 8: Evolving amplitude of the cyclical component in US real GDP (logarithms) for $n = 1$ with least informative prior on λ_c.

The HPD (Highest Posterior Density) regions for the estimated series are shown as well. These are computed by taking the 2.5 and 97.5 percentiles,

and the series of HPD regions form confidence bands that give the Bayesian counterpart of classical 95% confidence intervals. The smoothed cyclical component incorporates the prior information on periodicity. More generally, the Bayesian smoother accounts for the uncertainty in the parameter values - this contrasts with the classical case where parameters are assumed fixed at their estimates.

Posterior results for the cyclical parameters in the second order case, with the least informative prior on λ_c, are presented in figure 9. With the same prior as before, the marginal posterior of ρ now peaks around 0.71. The marginal posterior of $2\pi/\lambda_c$ peaks at just below thirty quarters, and there is significant mass above an eight-year period. The distribution of the period may conflict with expectations about the business cycle range. Noting the rather low estimate of σ_ζ^2 reported in table 1, it seems that the cycle is accounting for (low frequency) movements that are more plausibly attributed to variations in the trend.

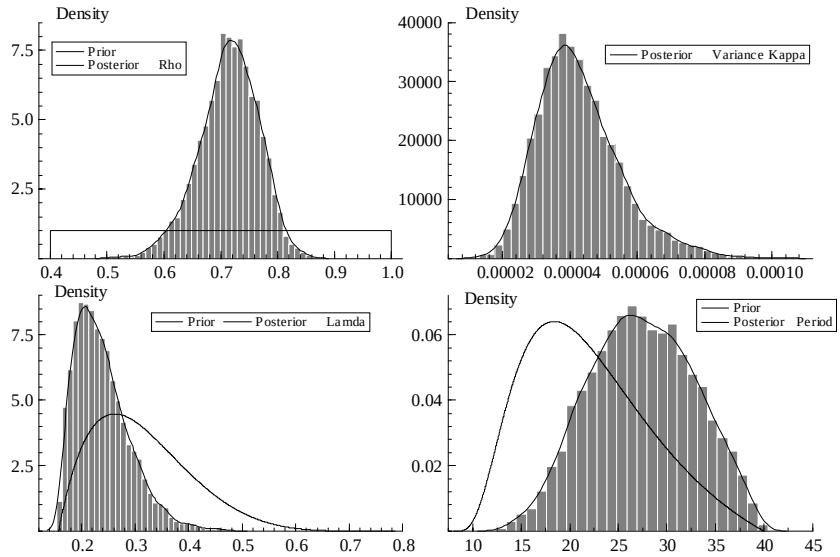

Figure 9: Marginal posterior densities of $\rho, \sigma_\kappa^2, \lambda_c$, and $2\pi/\lambda_c$ for $n = 2$ with least informative prior on λ_c for quarterly US real GDP (logarithms) from 1947:1 to 2001:4.

Table 1 shows posterior means and marginal likelihood estimates for $n = 1$ to 4, with the least informative prior on the frequency. When $n = 3$, there again seems to be some difficulty in estimating a reasonable cyclical, albeit

to a lesser extent than for the second order case. For $n = 4$, the results are slightly better, but the average period is still about one year above the value of five years obtained for the first order cycle.

Series	n	σ_ζ^2	σ_κ^2	σ_ε^2	ρ	λ_c	$2\pi/\lambda_c$	σ_ψ^2	q	$m(y)$
Real GDP	1	16.4	610	4	0.902	0.322	20.4	3638	0.00609	682.8
	2	4.65	435	102	0.715	0.239	27.4	6215	0.00118	686.4
	3	8.39	233	154	0.587	0.256	25.8	5545	0.00254	685.5
	4	15.2	171	165	0.486	0.273	24.5	4820	0.00631	684.1
Investment	1	11.9	22,487	24.2	0.880	0.289	22.5	10,436	0.000132	309.1
	2	5.03	11,124	4817	0.690	0.264	24.8	12,119	0.000049	309.8
	3	5.92	5295	6513	0.597	0.272	24.3	13,426	0.000056	311.7
	4	7.12	3086	7047	0.534	0.283	23.5	15,535	0.000068	311.9

Table 1: Posterior means for quarterly US real GDP from 1947:1 to 2001:4 for different values of n with a relatively noninformative prior on λ_c. This prior is based on a beta density $\beta(r, s)$ with $r = 2$. $2\pi/\lambda_c$ is the period in quarters, and σ_ψ^2 is the variance of the cycle. The signal to noise ratio $q = \sigma_\zeta^2/(\sigma_\psi^2 + \sigma_\varepsilon^2)$. The logarithm of the marginal likelihood $m(y)$ was estimated for each model using a LaPlace approximation. All variance parameters are multiplied by 10^7.

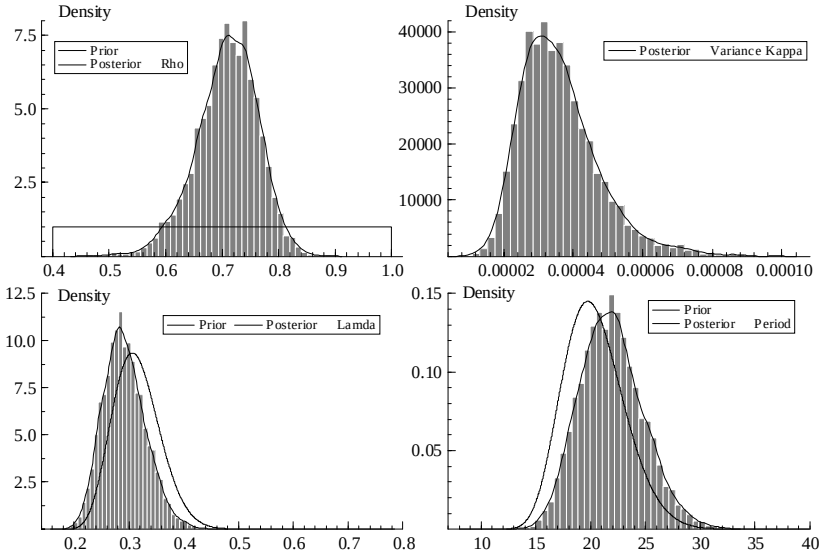

Figure 10: Marginal posterior densities of $\rho, \sigma_\kappa^2, \lambda_c$, and $2\pi/\lambda_c$ for $n = 2$, with intermediate prior on λ_c for quarterly US real GDP (logarithms) from 1947:1 to 2001:4.

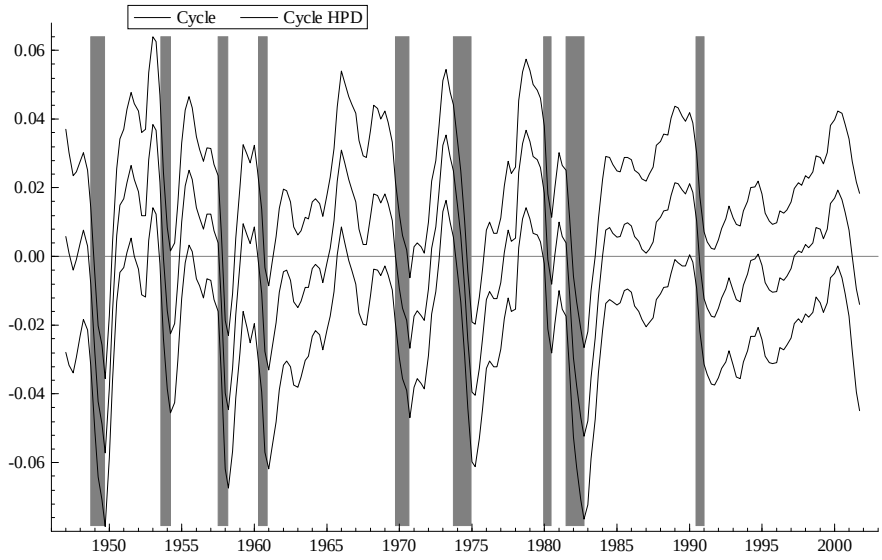

Figure 11: Estimated cycle in quarterly US real GDP (logarithms) and 95% confidence bands for $n = 2$, with intermediate informative prior on λ_c.

Two types of sharper priors, which were shown earlier in figure 3, are therefore investigated. These specify at the outset that the central period of the cycle focuses more narrowly around five years. The posterior means are recorded, for $n = 2$ and 3, in table 2. Note how increasing the sharpness of the prior leads to a decline in σ_κ^2 and an associated rise in σ_ζ^2. The trend becomes more flexible as the cycle concentrates more on mid-range frequencies.

The marginal posteriors for $n = 2$ are shown in figure 10 for the intermediate prior. The estimated cycle is shown in figure 11. The main difference with the first order cycle is its relative smoothness; this makes it easier to track movements in the business cycle. The higher value of σ_ε^2 in the second-order case means that a good deal of high frequency movement is consigned to the irregular.

Prior	n	σ_ζ^2	σ_κ^2	σ_ε^2	ρ	λ_c	$2\pi/\lambda_c$	σ_ψ^2	q	$m(y)$
Intermediate	2	8.48	360	111	0.709	0.292	21.9	4939	0.00248	685.2
	3	11.3	154	162	0.613	0.303	21.1	4594	0.00398	684.7
Sharp	2	10.4	333	115	0.706	0.312	20.2	4376	0.00315	684.7
	3	13.0	186	159	0.588	0.313	20.1	4432	0.00435	684.5

Table 2: Posterior means for quarterly US real GDP from 1947:1 to 2001:4 for $n = 2$ and 3 with more informative priors on λ_c. The intermediate prior was based on a $\beta(r, s)$ with $r = 10$, while for the sharp prior, $r = 100$. $2\pi/\lambda_c$ is the period in quarters, and σ_ψ^2 is the variance of the cycle. The signal to noise ratio $q = \sigma_\zeta^2/(\sigma_\psi^2 + \sigma_\varepsilon^2)$. $m(y)$ is the estimated logarithm of the marginal likelihood. All variance parameters are multiplied by 10^7.

For the second order cycle, using an informative prior solves the problem of the irregular likelihood and a plausible value is obtained for the posterior mean of the period. The use of the intermediate prior on λ_c seems to be a good choice. The second order model is attractive as it extracts a relatively smooth cycle with a minimum of complexity, and its marginal likelihood, $m(y)$, exceeds those obtained for higher order cycles.

4.3 Annual data

Annual time series are available over a fairly long period of time and this allows one to investigate issues concerning long-term changes in the business cycle. The enormous swing from the beginning of the Great Depression to the end of World War II constitutes a much longer and more pronounced cycle than is found in post-war data. To allow for longer periods we therefore consider priors for λ_c that have a mean of $2\pi/10$; in other words the average period is taken to be ten rather than five years. In addition we assume that the variance of the cyclical disturbances from 1929 to 1946 is ten times what it is elsewhere. The state space model has no difficulty handling such an extension. (A similar device could have been adopted for the frequency, that is the period from 1929 to 1946 could have been assumed to be double what it is elsewhere). The posterior results for $n = 1$ with the least informative prior on λ_c are shown in figure 12. The density of the damping factor peaks at a higher value than for the quarterly data, indicating more persistence in the annual cycle.

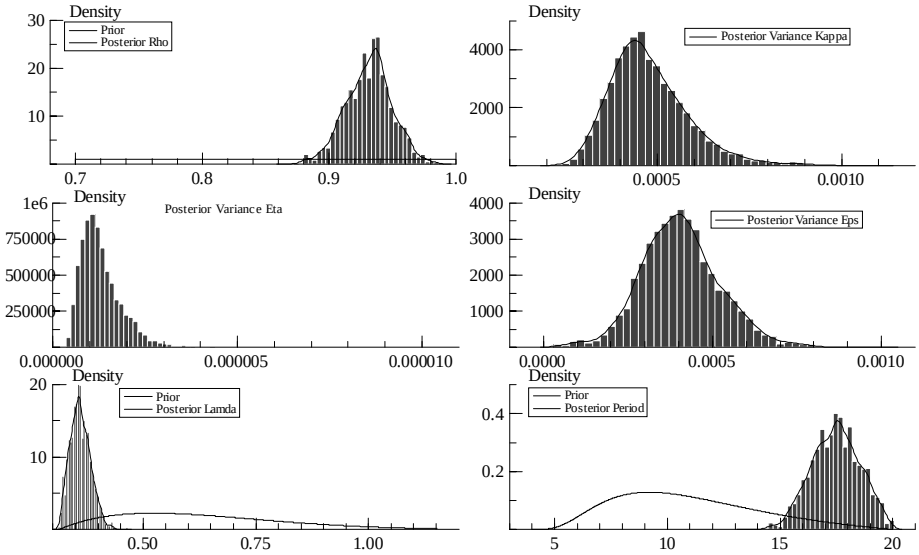

Figure 12: Marginal posterior densities of parameters with least informative prior on λ_c for annual US real GDP (logarithms) from 1870 to 1998.

Despite the fact that the prior expectations of periodicity centre around 10 years, the posterior of $2\pi/\lambda_c$ concentrates around 17.5 years, owing to the dominating influence of the long trade cycle swing from 1929 to 1946. Two sharper priors are considered, and the corresponding posterior densities for the frequency and period are shown in figure 13. It may be preferable to use the sharpest prior (shown on the right in the figure) to insure that a plausible range for $2\pi/\lambda_c$, corresponding to business cycle periods, is obtained. The resulting trend and cycle in this case are shown in figures 14 and 15.

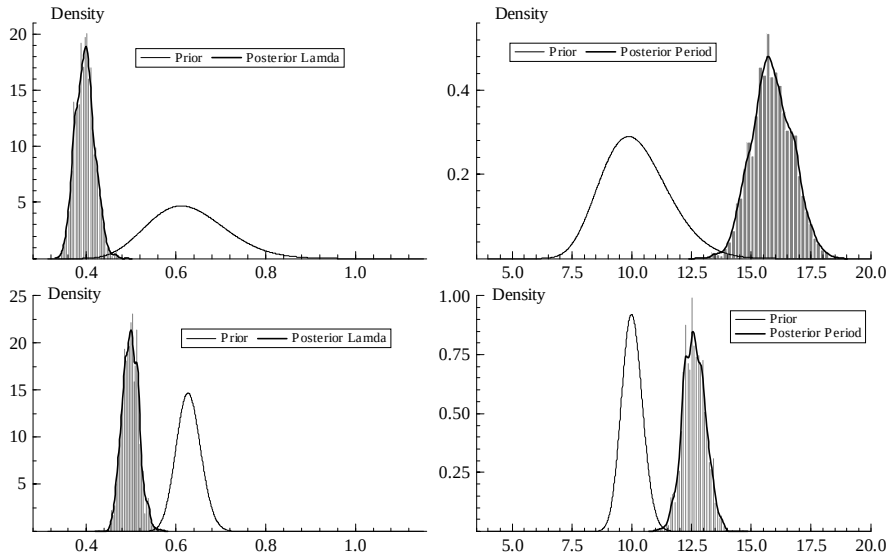

Figure 13: Marginal posterior densities of frequency and period with intermediate and very informative priors on λ_c for annual US real GDP (logarithms) from 1870 to 1998.

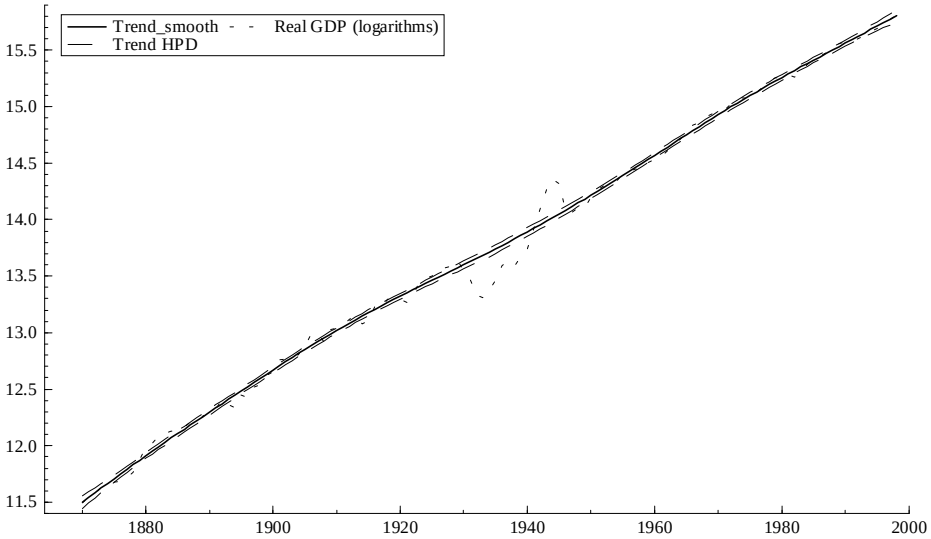

Figure 14: Estimated trend in annual US real GDP (logarithms) from 1870 to 1998 for $n = 1$ with most informative prior on λ_c.

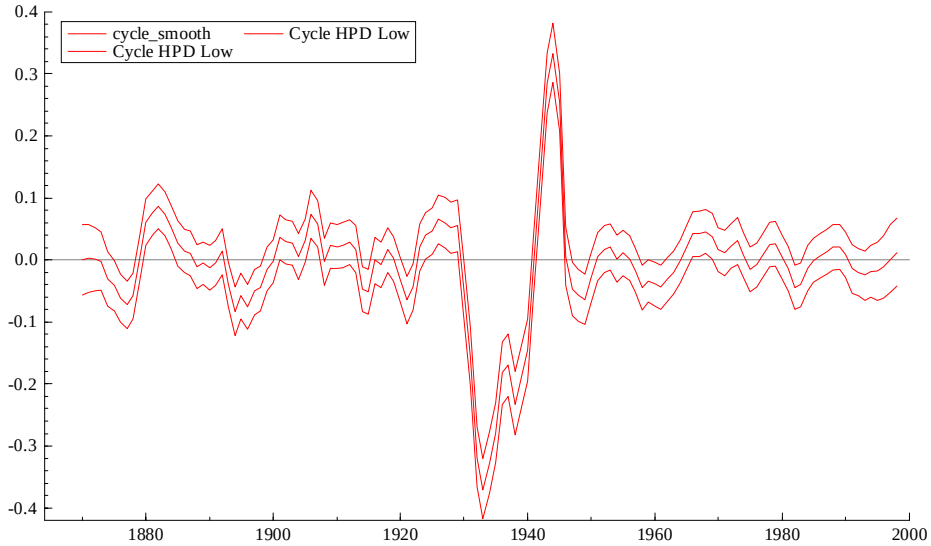

Figure 15: Estimated cycle in annual US real GDP (logarithms) from 1870 to 1998 for $n = 1$ with most informative prior on λ_c.

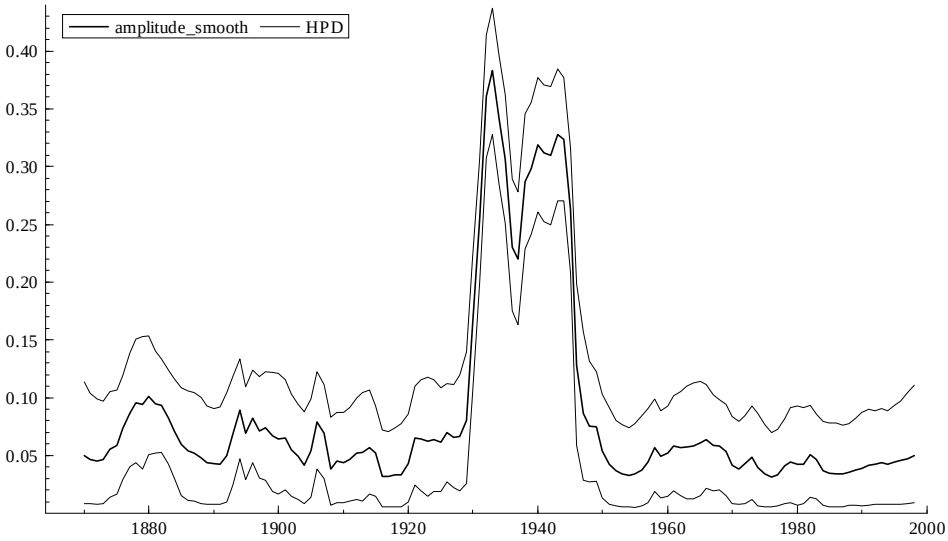

Figure 16: Evolving amplitude of the cyclical component in annual US real GDP (logarithms) from 1870 to 1998 for $n = 1$ with a very informative prior on λ_c.

An interesting question is whether business cycles have generally declined

in intensity since WWII. Figure 16 displays the evolving amplitude. The graph is dominated by the Great Depression and WWII. However, there is some indication that the average amplitude of the cycle is smaller in the post war period as compared with the period before 1929.

As in the case of quarterly real GDP, there is some difficulty in estimating plausible cycles in the annual series for higher order models; with noninformative priors the cyclical variance falls to near zero so that the trend accounts for most of the variation in the series. Therefore, for $n > 1$ more informative priors on σ_ζ^2 were used to ensure a minimal degree of smoothness in the trend. The estimated cyclical component for $n = 2$ is displayed in figure 17. Note the increased smoothness relative to the first order case. Posterior means for different values of n are shown in table 4 for the sharp prior; with less informative priors on λ_c, the posterior mass of the period shifts to higher values while the results for the other parameters remain similar. In the case $n = 4$, a substantial amount of noise is removed from the cycle, as can be seen from the higher estimate for σ_ε^2. The turning points in the cyclical component become even more apparent; see figure 18.

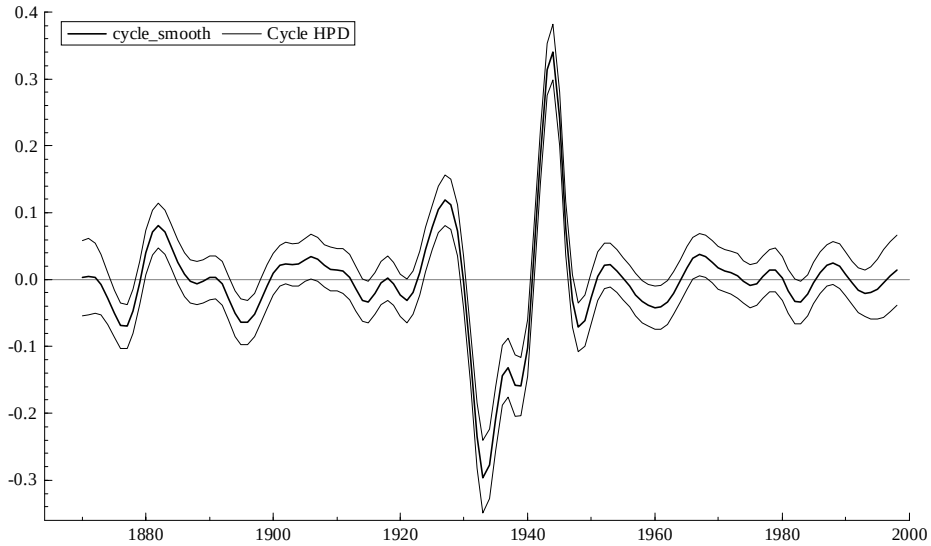

Figure 17: Estimated cycle in annual US real GDP (logarithms) from 1870 to 1998 for $n = 2$ with a very informative prior on λ_c.

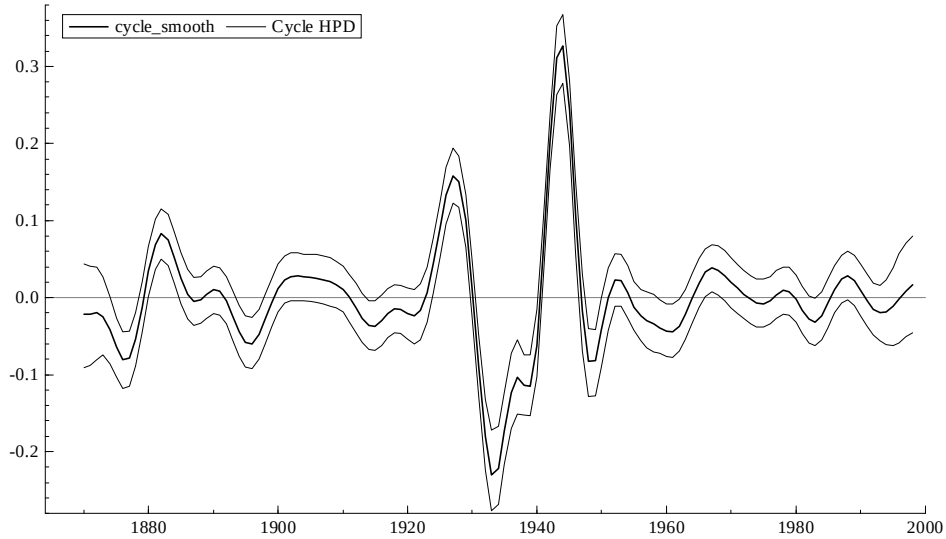

Figure 18: Estimated cycle in annual US real GDP (logarithms) from 1870 to 1998 for $n = 4$ with a very informative prior on λ_c.

Order n	σ^2_ζ	σ^2_κ	σ^2_ε	ρ	λ_c	$2\pi/\lambda_c$
1	60.7	5416	3413	0.918	0.507	12.4
2	18.8	417	7847	0.897	0.525	11.9
3	20.7	64.4	8296	0.872	0.549	11.5
4	12.9	3.11	10,469	0.883	0.542	11.6

Table 3: Posterior means for annual US real GDP from 1870 to 1998 for different values of n with an informative prior on λ_c, centred at $2\pi/10$. For $n = 2$, 3, the shape and scale of the inverted gamma prior on σ^2_ζ were set to 20 and 2 x 10^{-5}, respectively. For $n = 4$, the shape and scale of the inverted gamma prior on σ^2_ζ were set to 100 and 10^{-4}, and additionally, a moderately informative prior on σ^2_ε was used, with shape and scale equal to 10 and 10^{-5}. The period in years is $2\pi/\lambda_c$. All variance parameters are multiplied by 10^7.

4.4 Forecasts

In making predictions about future cyclical and trend movements, the Bayesian approach accounts for uncertainty in the hyperparameters, just as it does in smoothing.

The forecast function $E(y_{t+h}|Y_T, \theta)$ in (10) may be expressed in terms of trend and cycle components. The forecast of the trend is simply a deterministic trend with the slope and level given by the Kalman filter estimates at time T. The forecast function for the cyclical component with $n = 1$ is a damped cycle

$$\widetilde{\psi}_{1,T+h|T}(\theta) = \rho^h \left(\psi_{1,T} \cos[h\lambda_c] + \psi^*_{1,T} \sin[h\lambda_c]\right)$$

where h is the forecast horizon and $\psi_{1,T}, \psi^*_{1,T}$ are the Kalman filter estimates at the end of the sample. For the second-order cyclical,

$$\begin{aligned}\widetilde{\psi}_{2,T+h|T}(\theta) &= \left(h\rho^{h-1}\right)\left(\psi_{1,T} \cos[(h-1)\lambda_c] + \psi^*_{1,T} \sin[(h-1)\lambda_c]\right) \\ &\quad +\rho^h \left(\psi_{2,T} \cos[h\lambda_c] + \psi^*_{2,T} \sin[h\lambda_c]\right)\end{aligned}$$

The combination of the two levels of the cycle means that it is possible for the amplitude of the predicted cycle to increase before it eventually damps down to zero.

The Bayesian predictions at horizon h are obtained by averaging over the J draws. Thus for the cycle

$$\widehat{\psi}_{n,T+h|T} = \frac{1}{J}\sum_{j=1}^{J} \widetilde{\psi}_{n,T+h|T}(\theta^{(j)})$$

Given the structure of the model, there is a straightforward way to assess the forecast uncertainty, which avoids the numerically demanding task of estimating the predictive density. The variance of the h-step ahead prediction may be decomposed as

$$var(\psi_{n,T+h|T}|Y) = \int var[\psi_{n,T+h|T}(\theta)|Y, \theta]p(\theta|Y)d\theta + var[\widetilde{\psi}_{n,T+h|T}(\theta)] \quad (11)$$

The second variance on the right hand side is taken with respect to the posterior and so may be estimated by

$$var[\widetilde{\psi}_{n,T+h|T}(\theta)] = \left(\frac{1}{J}\sum \left[\widetilde{\psi}_{n,T+h|T}(\theta^{(j)})\right]^2\right) - \left(\widehat{\psi}_{n,T+h|T}\right)^2$$

The first term on the right hand side of (11) is equal to the mean of the classical forecast variance. Since the model in (5) may be cast in state space form, the forecasting routine in the SsfPack suite of programs may be used; see the appendix for details. In simple cases, an analytical expression for $var[\psi_{n,T+h|T}(\theta)|Y,\theta]$ may be obtained.

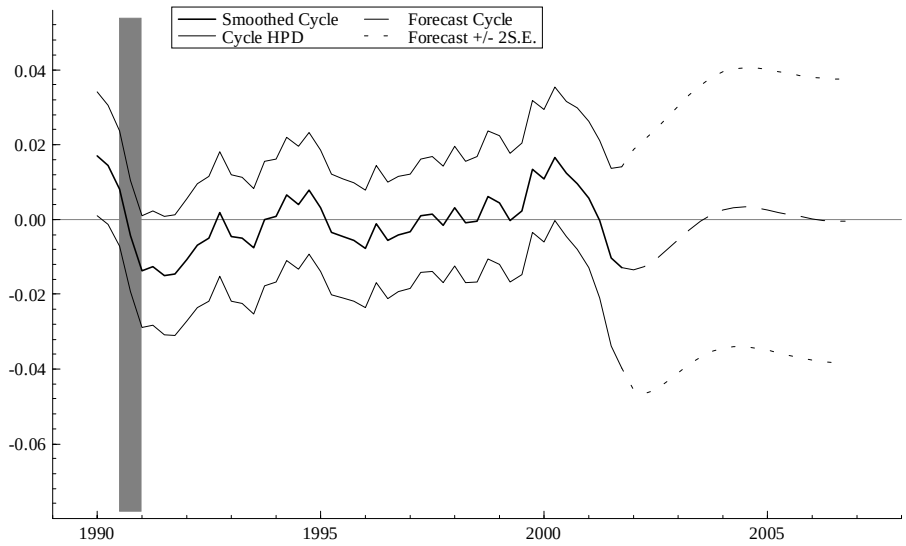

Figure 19: Forecast of cyclical component in US real GDP (logarithms) for $n = 1$ with least informative prior on λ_c.

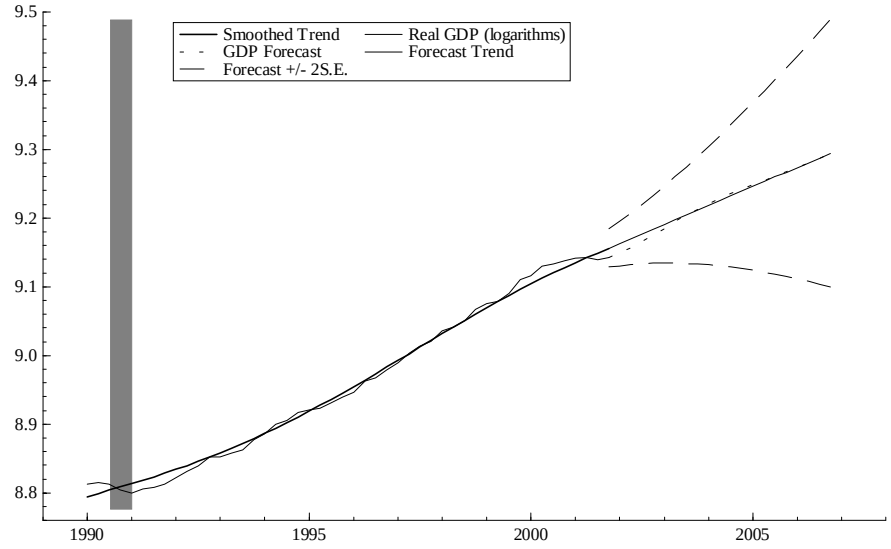

Figure 20: Forecast US real GDP (logarithms) for $n = 1$ with least informative prior on λ_c.

The forecast of the cyclical component, that is the time series of posterior means, along with bands lying at plus or minus two standard deviations, are shown in figure 19 for up to 20 quarters ahead. These confidence bands for the predictions correspond roughly to the 95% HPDs of the smoothed cyclical component, but to the extent that the predictive densities depart from normality, the two measures will not be equivalent. The forecasts show a cyclical recovery lasting until the early part of 2004; the standard error of the prediction gradually widens as the horizon is extended. The corresponding forecasts for the real GDP series and the forecast MSE are shown in figure 20. There are some differences in cyclical forecasts for higher n. In figure 21 based on the second order component, the recovery is anticipated to occur sooner and to be more moderate.

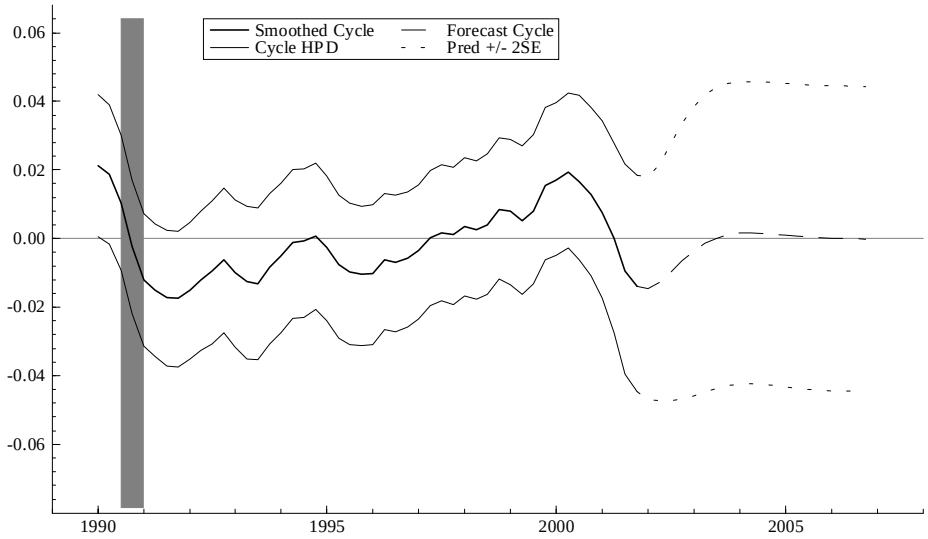

Figure 21: Forecast of cyclical component in US real GDP (logarithms) for $n = 2$ with informative prior on λ_c

5 Multivariate model

Modeling several series jointly can produce more precise parameter estimates by pooling information. A multivariate UC model may be set up for N series as

$$y_t = \mu_{m,t} + \psi_{n,t} + \varepsilon_t, \qquad \varepsilon_t \sim NID(0, \Sigma_\varepsilon), \quad t = 1, ..., T, \tag{12}$$

where the $N \times 1$ vector $y_t = (\ y_t^1, ..., \ y_t^N)'$ and similarly for $\mu_{m,t}, \psi_{n,t}$ and $\boldsymbol{\varepsilon}_t$. Σ_ε is an $N \times N$ positive semi-definite matrix. For $m = 2$, the trend is

$$\begin{aligned} \mu_t &= \mu_{t-1} + \beta_{t-1}, \\ \beta_t &= \beta_{t-1} + \zeta_t, \qquad \zeta_t \sim NID(0, \Sigma_\zeta), \end{aligned} \tag{13}$$

where $\boldsymbol{\Sigma}_\zeta$ is an $N \times N$ positive semi-definite matrix. In the notation of (2) and (3) μ_t, the level, is μ_{2t}, while β_t, the slope, is μ_{1t}.

The *similar cycle* model, introduced by Harvey and Koopman (1997) and implemented in the STAMP package of Koopman et al (2000), is defined for

$n = 1$ as

$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \left[\rho \begin{pmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{pmatrix} \otimes I_N \right] \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix}, \quad t = 1, ..., T, \tag{14}$$

where κ_t and κ_t^* are $N \times 1$ vectors of the disturbances such that

$$E(\kappa_t \kappa_t') = E(\kappa_t^* \kappa_t^{*'}) = \Sigma_\kappa, \quad E(\kappa_t \kappa_t^{*'}) = 0, \tag{15}$$

and $\mathbf{\Sigma}_\kappa$ is an $N \times N$ covariance matrix. The model may be readily extended to higher order cycles as in Harvey and Trimbur (2002) and appendix D. Because the damping factor and the frequency, ρ and λ_c, are the same in all series, the cycles in the different series have similar properties; in particular their movements are centred around the same period. This seems eminently reasonable if the cyclical movements all arise from a similar source such as an underlying business cycle. Furthermore, the restriction means that it is often easier to separate out trend and cycle movements when several series are jointly estimated.

The Bayesian analysis extends directly to this multivariate model. The essence of the MCMC routine remains the same. The details are in appendix D. We illustrate the method with a bivariate model for real GDP and investment. Table 4 shows results for the least informative prior on λ_c with an essentially noninformative inverted Wishart prior on the variance matrix parameters.

Series: US real GDP			
n	σ^2_ζ	σ^2_κ	σ^2_ε
1	17.4	643	22
2	17.5	470	81
3	17.0	228	153
4	21.8	131	177

Series: US Investment			
n	σ^2_ζ	σ^2_κ	σ^2_ε
1	32.6	22,818	23
2	10.0	12,799	4383
3	6.23	5977	6472
4	10.1	3336	7119

Correlations and cyclical parameters

n	υ_ζ	υ_κ	υ_ε	ρ	$2\pi/\lambda_c$
1	0.847	0.811	0.255	0.876	23.4
2	0.574	0.909	0.459	0.670	25.7
3	0.303	0.920	0.642	0.581	24.9
4	0.550	0.935	0.660	0.513	24.1

Table 4: Posterior means for a bivariate model for quarterly US real GDP and investment from 1947:1 to 2001:4 with a relatively noninformative prior on λ_c. The priors on the variance matrices are essentially noninformative inverted Wishart densities. All variance parameters are multiplied by 10^7. The parameter υ_ζ denotes the correlation between the slope disturbances in the two series, and ρ and λ_c are the shared cyclical parameters.

Posterior results for the cyclical parameters with $n = 2$ are shown in figure 22. The draws for the variance matrices may be directly used to form draws for the correlations between the components of the series. Thus, the posterior density of the cyclical error correlation peaks at around 0.9, indicating significant co-movements in the cycles. The estimated common period of oscillation is about two quarters less than the posterior mean for the univariate model for real GDP with the least informative prior.

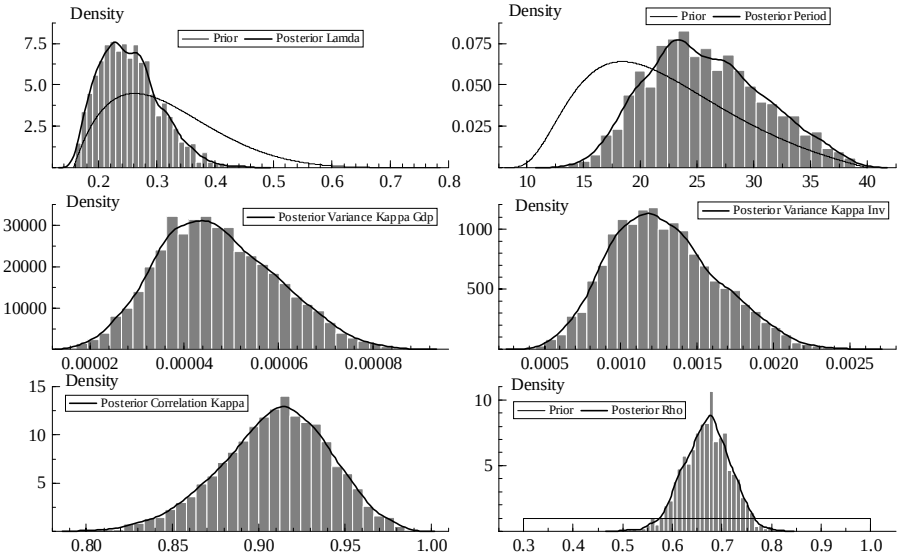

Figure 22: Marginal posterior densities for the cyclical parameters in a bivariate model of quarterly US real GDP and investment (logarithms) from 1947:1 to 2001:4. The parameter correlation kappa denotes the correlation between the cyclical disturbances in the two series.

Figure 23 shows the cycle produced by the bivariate model for $n = 2$. The overall shape resembles that of figure 11, but the confidence bands are at times narrower. Also, a steeper upswing in the cycle in the 1990s is apparent in figure 23. The cycles estimated from the bivariate models are often more pronounced; for instance, for real GDP the cyclical variance rises by around 20% when $n = 2$.

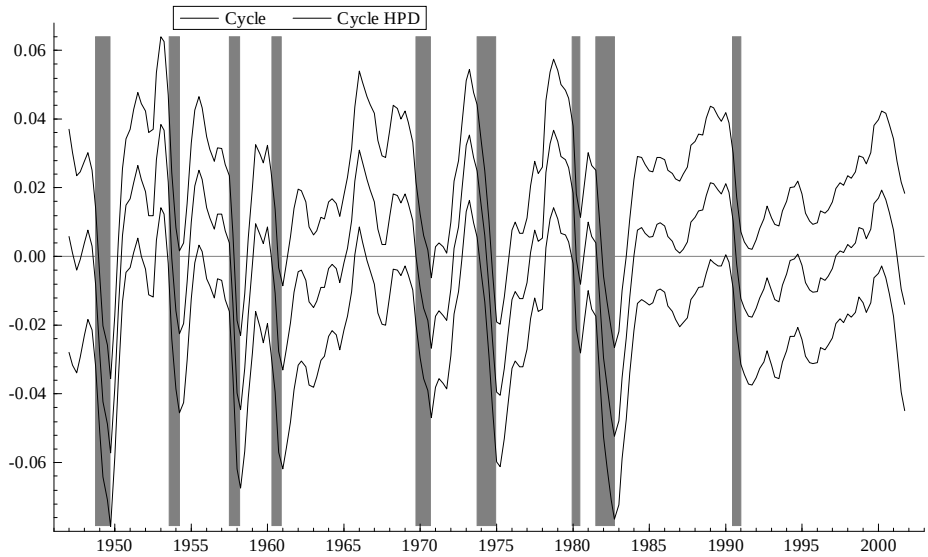

Figure 23: Estimated cycle in quarterly US real GDP (logarithms) from 1947:1 to 2001:4, based on a bivariate model with investment.

To the extent that somewhat more plausible business cycle periods result from combining the information in the real GDP and investment series, this approach provides another way to help address the problem of the likelihood surface in the case of univariate models for real GDP.

6 Conclusion

A structural time series model provides a framework for extracting trends and cycles. This article has investigated the Bayesian treatment of such a model, paying particular attention to the cyclical component and the way in which prior information on periodicity can be used. Markov chain Monte Carlo methods are successfully adapted to deal with the models, including those with higher order cycles of the kind introduced recently by Harvey and Trimbur (2002), and those in which two series are considered jointly.

The approach is illustrated with an application to US macroeconomic time series, where smooth cycles are successfully extracted from real GDP and investment. These cycles have a simple interpretation in terms of the percentage by which they exceed or fall below the long-term level and they are consistent with the NBER dating of business cycles. The approach is

shown to be viable in situations where maximum likelihood estimation failed to yield plausible results.

ACKNOWLEDGMENTS

Earlier versions of this paper was presented at the Cambridge PhD Workshop in Econometrics and at the NAKE conference at the Netherlands Central Bank. Trimbur wishes to thank the Cambridge Commonwealth Trust and the Richard Kahn Fund for financial support, and is grateful to the Tinbergen Institute for its hospitality and financial support in autumn 2001. Harvey thanks the Economic and Social Research Council (ESRC) for support as part of a project on Dynamic Common Factor Models for Regional Time Series, grant number L138 25 1008.

References

Baxter, M. and King R. G., 1999. Measuring business cycles: approximate band-pass filters for economic time series, Review of Economics and Statistics 81, 575-93.

Bos, C., R. Mahieu, and van Dijk, H.K., 2000. Daily exchange rate behaviour and hedging of currency risk, Journal of Applied Econometrics 15, 671-96.

Carter, C. K. and Kohn R., On Gibbs sampling for state space models, Biometrika 81, 541-53.

Chib, S. and Greenberg E., 1996. Markov chain Monte Carlo simulation methods in econometrics, Econometric Theory 12, 409-31.

Cogley, T. and Nason J.M., 1995. Effects of the Hodrick-Prescott filter on trend and difference stationary time series: implications for business cycle research, Journal of Economic Dynamics and Control 19, 253-78.

DiCiccio, T., Kass R., Raftery A., and Wasserman L., 1997. Computing Bayes factors by combining simulation and asymptotic approximations, Journal of the American Statistical Association 92, 903-15.

Doornik, J. A., 1999. Ox: An Object-Oriented Matrix Programming Language (London: Timberlake Consultants Ltd.).

Durbin, J., and Koopman S.J., 2002. A simple and efficient simulation smoother, Biometrika 89, 603-16.

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

Gomez, V., July 2001. The use of Butterworth filters for trend and cycle estimation in economic time series, Journal of Business and Economic Statistics 19, 365-73.

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

Harvey, A. C., and Jaeger A., 1993. Detrending, stylised facts and the business cycle, Journal of Applied Econometrics 8, 231-47.

Harvey, A. C., and Koopman S.J., 1997. Multivariate structural time series models. In C. Heij et al.(eds), System dynamics in economic and financial models, Chichester: Wiley and Sons.

Harvey, A. C., and Trimbur T.M., 2002. General model-based filters for extracting trends and cycles in economic time series, Review of Economics and Statistics, to appear. (http://www.econ.cam.ac.uk/faculty/harvey/papers.htm)

Hodrick, R. J. and Prescott E.C., 1997. Postwar US business cycles: an empirical investigation, Journal of Money, Credit and Banking 24, 1-16.

Kass, R. E. and Raftery A.E., 1995. Bayes factors, Journal of the American Statistical Association 24, 773-95.

Kitagawa, G., and Gersch W., 1996. Smoothness priors analysis of time series (Berlin: Springer-Verlag).

Kohn, R., C. Ansley, and Wong C.H., 1992. Nonparametric spline regression with autoregressive moving average errors, Biometrika 79, 335-46.

Koop, G. and van Dijk H.K., 2000. Testing for integration using evolving trend and seasonals models: a Bayesian approach, Journal of Econometrics 97, 261-91.

Koopman, S. J., Shephard N., and Doornik J., 1999. Statistical algorithms for models in state space using SsfPack 2.2, Econometrics Journal 2, 113-66.

Koopman, S. J., Harvey A.C., Doornik J.A., and Shephard N., 2000. STAMP 6.0 Structural Time Series Analysis Modeller and Predictor (London: Timberlake Consultants Ltd.).

Magnus, J. and Neudecker H., 1999. Matrix Differential Calculus with Applications in Statistics and Econometrics (Chichester: John Wiley & Sons).

Murray, C. J., 2002. Cyclical properties of Baxter-King filtered time series, Review of Economics and Statistics, to appear.

Trimbur, T. M., 2002. Properties of a general class of stochastic cycles, mimeo.

Young, P., 1984. Recursive Estimation and Time Series Analysis (Berlin: Springer-Verlag).

A Priors

The priors on λ_c derive from the beta family of densities and are based on expectations of period lying in the business cycle range. A random variable X has a beta distribution with parameters R and S, that is $X \sim \beta(R, S)$ for $R, S > 0$, if

$$f(X) = \frac{\Gamma(R+S)}{\Gamma(R)\Gamma(S)} X^{R-1}(1-X)^{S-1}, \quad 0 < X < 1,$$

where $\Gamma(.)$ is the gamma function.

A beta density is defined on an arbitrary interval (a, b),where $a, b > 0$, by setting $Y = a + (b-a)X$. The mean and variance of the transformed variable are given by

$$\mu_\beta = a + (b-a)\frac{R/S}{R/S+1}, \qquad \sigma^2_\beta = (b-a)^2 \frac{RS}{(R+S)^2(R+S+1)}$$

The priors on λ_c were all constructed to have mean $\pi/10$ (5 year period) and to lie between $\pi/20$ (10 year period) and $\pi/4$ (2 year period). Therefore, $(5/\pi)(\lambda_c - \pi/20) \sim \beta(R, S)$. The reason for constraining λ_c to an interval within $[0,\ \pi]$ is that frequencies near zero and π are implausible. For the

wide prior, $R = 2$. For the intermediate and sharp cases, R was set to 10 and 100, respectively. S was then determined by the value of the mean and the boundaries.

The priors for the variance parameters $\sigma_\zeta^2, \sigma_\kappa^2, \sigma_\varepsilon^2$ are members of the inverted gamma family. The random variable X has an inverted gamma distribution with shape c and scale S, denoted $X \sim IG(c, S)$, if its density is given by

$$f(X) = \frac{1}{\Gamma(c/2)(2/S)^{c/2}}(X)^{-\frac{1}{2}(c+2)}e^{-S/(2X)}, \quad X > 0 \tag{17}$$

The parameters c and S determine the location and spread of the distribution and the mean and variance σ^2 are given by

$$\mu = S/(c-2),\ c > 2, \qquad \sigma^2 = 2\mu^2/(c-4),\ \ c > 4$$

Low values of c correspond to less informative priors. The mean and variance of $1/X$ are given by c/S and $2c/S^2$ respectively.

B SSF and initial conditions

This appendix presents the state space form of the class of models in (1). The density of the states conditional on the hyperparameters, which plays a key role in the MCMC routine presented in the next appendix, is then derived.

The model defined in equations (1) to (5) expresses the observed process as a sum of unobserved components. The $(m+2n) \times 1$ state vector contains the trend, cycle, and the associated processes used in their definition, that is

$$\begin{aligned} \alpha_t &= (\ \mu_{m,t}\ \ \mu_{m-1,t}\ \ .\ \ .\ \ .\ \ \mu_{1,t}\ \ \psi_{n,t}\ \ \psi^*_{n,t}\ \ \psi_{n-1,t}\ \ .\ \ .\ \ .\ \ \psi_{1,t}\ \ \psi^*_{1,t}\ )' \\ &= (\mu_t', \psi_t') \end{aligned} \tag{19}$$

Denote by α the $(m + 2n)$ x T matrix formed by stacking the complete set of state vectors over the sample period:

$$\alpha = [\alpha_1, ..., \alpha_T] \tag{20}$$

This will be referred to as the state matrix.

The measurement equation is

$$y_t = z_t \alpha_t + \varepsilon_t \tag{21}$$

where the vector z_t contains ones in the first and $(m+1)-th$ positions and zeroes elsewhere.

The transition equation for the trend is

$$\mu_t = (I_m + S_m)\mu_{t-1} + c_m \zeta_t \tag{22}$$

where S_m is a matrix with ones on the diagonal strip to the right of the main diagonal and zeros elsewhere. That is, the row i, column $i+1$ element of S_m equals 1 for $i = 1, ..., m-1$, and all other elements equal 0. The $m \times 1$ column c_m has a one in the last position and zeroes elsewhere. The covariance matrix of the disturbance vector is therefore

$$Var(c_m \zeta_t) = \sigma_\zeta^2 c_m c_m' \tag{23}$$

The transition equation for the cyclical part of the state vector is

$$\psi_t = T_\psi \psi_{t-1} + c_n \otimes \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix} \tag{24}$$

where

$$T_\psi = I_n \otimes \begin{bmatrix} \rho \cos \lambda_c & \rho \sin \lambda_c \\ -\rho \sin \lambda_c & \rho \cos \lambda_c \end{bmatrix} + S_n \otimes I_2 \tag{25}$$

The covariance matrix of the disturbance vector is

$$Var \left\{ c_n \otimes \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix} \right\}) = c_n c_n' \otimes \begin{bmatrix} \sigma_\kappa^2 & 0 \\ 0 & \sigma_\kappa^2 \end{bmatrix}$$

The Kalman filter is applied by initialising the nonstationary trend vector, μ_t, with a diffuse prior. The unconditional distribution of the cyclical state vector ψ_t has mean zero, and the general form of the covariance matrix for the nth order cyclical is

$$\Gamma = \begin{bmatrix} \Gamma_{n,n} & \Gamma_{n,n-1} & ... & \Gamma_{n,1} \\ \Gamma_{n-1,n} & ... & ... & \Gamma_{n-1,1} \\ ... & ... & ... & ... \\ \Gamma_{1,n} & \Gamma_{1,n-1} & ... & \Gamma_{1,1} \end{bmatrix} \tag{26}$$

with the 2 x 2 block in position i, j given by

$$\Gamma_{i,j} = \frac{\sigma_\kappa^2 \sum_{r=0}^{i-1} \begin{pmatrix} i-1 \\ r \end{pmatrix} \begin{pmatrix} j-1 \\ r+j-i \end{pmatrix} \rho^{2r}}{(1-\rho^2)^{i+j-1}} T^{j-i},$$

for all $1 \leq i, j \leq n$ such that $i \leq j$; see Trimbur (2002) for the proof and further details. Note that the block index goes from n at the top (left) to 1 at the bottom (right) to remain consistent with the definition of the cyclical state vector. The upper triangular part is given by the transpose, that is $\Gamma_{i,j} = \Gamma'_{j,i}$ for $j \leq i$.

The variance of the cycle is obtained from the first element in the top-left block, $\Gamma_{n,n}(\rho)$, and is given by

$$var(\psi_n) = \frac{\sigma_\kappa^2 \sum_{i=0}^{n-1} \begin{pmatrix} n-1 \\ i \end{pmatrix}^2 \rho^{2i}}{(1-\rho^2)^{2n-1}} \tag{28}$$

The conditional density of the state matrix, given the parameters, is an integral part of the conditional posteriors in the Gibbs sampler:

$$p(\alpha|\theta) = p(\alpha_1|\theta) \prod_{t=2}^{T} p(\alpha_t|\alpha_{t-1}, \theta) \tag{29}$$

As noted previously, normality is assumed throughout the analysis. The density of the initial state depends only on the cyclical parameters since the trend at time $t = 1$ is diffuse:

$$p(\alpha_1|\sigma_\kappa^2, \rho, \lambda_c) \propto |\Gamma|^{-1/2} \exp(-\frac{1}{2} {\psi_1}' \Gamma^{-1} \psi_1)$$

The vector ψ_1 contains the starting values of the cyclical elements in the state vector. Conditional on α_{t-1}, some elements of α_t are known so we define the reduced state vector, transition matrix, and covariance matrix by:

$$\alpha_t^* = \begin{bmatrix} \mu_{1,t} \\ \psi_{1,t} \\ \psi_{1,t}^* \end{bmatrix}, \quad T^* = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \rho\cos\lambda_c & \rho\sin\lambda_c \\ 0 & -\rho\sin\lambda_c & \rho\cos\lambda_c \end{bmatrix}, \quad \Omega^* = \begin{bmatrix} \sigma_\zeta^2 & 0 & 0 \\ 0 & \sigma_\kappa^2 & 0 \\ 0 & 0 & \sigma_\kappa^2 \end{bmatrix}$$

Then the one-step ahead density is:

$$p(\alpha_t|\alpha_{t-1},\theta) = |\Omega^*|^{-1/2} \exp\left\{-\frac{1}{2}(\alpha_t^* - T^*\alpha_{t-1}^*)'\Omega^{*-1}(\alpha_t^* - T^*\alpha_{t-1}^*)\right\}$$

Therefore,

$$p(\alpha|\theta) \propto |\Gamma|^{-1/2} |\Omega^*|^{-(T-1)/2} \exp\left\{-\frac{1}{2}\psi_1'\Gamma^{-1}\psi_1 - \frac{1}{2}\sum_{t=2}^{T}(\alpha_t^* - T^*\alpha_{t-1}^*)'\Omega^{*-1}(\alpha_t^* - T^*\alpha_{t-1}^*)\right\}$$

Given the partitioned structure of T^*, this can be rewritten as

$$\begin{aligned} p(\alpha|\theta) \quad &\propto \quad |\Gamma|^{-1/2} \sigma_\kappa^{-2(T-1)} \sigma_\zeta^{-(T-1)} \exp\left\{-\frac{1}{2}\psi_1'\Gamma^{-1}\psi_1 - \frac{1}{2\sigma_\kappa^2}\sum_{t=2}^{T} c_t\right\} \quad (30) \\ &\times \exp\left\{-\frac{1}{2\sigma_\zeta^2}\sum_{t=2}^{T}(\mu_{1,t} - \mu_{1,t-1})^2\right\} \end{aligned}$$

where

$$c_t = (\psi_{1,t} - \rho\cos\lambda_c\psi_{1,t-1} - \rho\sin\lambda_c\psi_{1,t-1}^*)^2 + (\psi_{1,t}^* + \rho\sin\lambda_c\psi_{1,t-1} - \rho\cos\lambda_c\psi_{1,t-1}^*)^2$$

C MCMC routine

This section sets out the details of the Gibbs sampler used to obtain posterior draws and component estimates. The steps of the MCMC routine are outlined, and the method for drawing from each conditional posterior distribution is discussed. Details on simulation output are then noted.

In designing an efficient MCMC routine, it is convenient to produce draws $\{\theta^{(i)},\ \alpha^{(i)}\}$ from the joint posterior $p(\theta, \alpha|Y)$. In this case, the complete conditional densities of the parameters are known up to proportionality, and can be handled using standard methods. Similar algorithms, which capitalise on the state space framework, may be found in Carter and Kohn (1994) and Koop and van Dijk (2000).

The four sections of the routine correspond to a set of complete conditional densities for the joint posterior density $p(\theta, \alpha|Y)$. Taken on their own, the variates $\{\theta^{(i)}\}$ represent draws from the marginal posterior of the parameter vector. Separately, the $\{\alpha^{(i)}\}$ serve as draws from $p(\alpha|Y)$.

We start with an initial value for the parameter vector, $\theta^{(0)}$. Each iteration i involves the following sequence of draws, organised into four steps:

1. $\alpha^{(i)}$ is drawn from $p(\alpha|\theta^{(i)}, Y)$:

The state space form of the model was shown in the previous appendix. This enables the direct application of the general simulation smoother of Durbin and Koopman (2002).

2. The variances $\{\sigma_\kappa^{2\,(i)}, \sigma_\zeta^{2\,(i)}, \sigma_\varepsilon^{2\,(i)}\}$ are sampled as a group from the joint density $p(\sigma_\kappa^2, \sigma_\zeta^2, \sigma_\varepsilon^2|\rho^{(i-1)}, \lambda_c^{(i-1)}, \alpha^{(i)}, Y)$:

As the priors on all parameters are assumed independent, we start with

$$p(\sigma_\kappa^2, \sigma_\zeta^2, \sigma_\varepsilon^2|\rho, \lambda_c, \alpha, Y) \;\propto p(Y|\alpha, \theta)p(\alpha|\theta)p(\sigma_\kappa^2)p(\sigma_\zeta^2)p(\sigma_\varepsilon^2) \tag{31}$$

Conditional on ρ and λ_c, the reduced state transition matrix T^* is fixed. The unconditional variance of the cyclical part of the state can be written as:

$$\Gamma = \sigma_\kappa^2 V(\rho, \lambda_c, n) \tag{32}$$

where V is a $2n$ x $2n$ matrix that depends only on ρ and λ_c. The determinant of Γ is therefore proportional to σ_κ^{4n}, and it follows that

$$p(\alpha|\theta) \propto \sigma_\kappa^{-2(T+n-1)} \exp\left\{-\frac{1}{2\sigma_\kappa^2}\psi_1{}'V^{-1}\psi_1 - \frac{1}{2\sigma_\kappa^2}\sum_{t=2}^{T} c_t\right\} \; \sigma_\zeta^{-(T-1)} \exp\left\{-\frac{1}{2\sigma_\zeta^2}\sum_{t=2}^{T}(\mu_{1,t} - \mu_{1,t-1}\right.$$

Furthermore,

$$p(Y|\alpha, \theta) \propto \sigma_\varepsilon^{-T} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\sum_{t=1}^{T}(y_t - z_t\alpha_t)^2\right\} \tag{33}$$

Given inverted gamma priors on the variances, the conditional posterior therefore factors into three independent inverted gamma densities. This is related to the d-inverse gamma class of models in Fruewirth-Schnatter (1994), but the models investigated here do not fall into this class given the presence of the cyclical component. The use of this natural conjugate framework on a

conditional basis simplifies the MCMC routine. It also remains flexible since a broad range of prior shapes for the variances are available by changing the shape and scale.

3. $\{\rho^{(i)}\}$ is drawn from $p(\rho|\sigma_\kappa^{2\,(i)}, \sigma_\zeta^{2(i)}, \sigma_\varepsilon^{2(i)}, \lambda_c^{(i-1)}, \alpha^{(i)}, Y)$:

The parameter ρ determines how the white noise shocks feed through to the cyclical component over time. A uniform prior is used to ensure the value of ρ lie between zero and one, but otherwise gives no information about its location within the interval. Denote the prior by $p(\rho)$. Using Bayes theorem for densities,

$$p(\rho|\sigma_\kappa^2, \sigma_\zeta^2, \sigma_\varepsilon^2, \lambda_c, \alpha, Y) \propto p(\rho)p(Y|\alpha, \theta)p(\alpha|\theta) \propto p(\rho)p(\alpha|\theta)$$

$$p(\rho|\sigma_\kappa^2, \sigma_\zeta^2, \sigma_\varepsilon^2, \lambda_c, \alpha, Y) \propto p(\rho)\left|\Gamma\right|^{-1/2} \exp\left\{-\frac{1}{2}\psi_1{}'\Gamma^{-1}\psi_1 - \frac{1}{2\sigma_\kappa^2}\sum_{t=2}^{T} c_t\right\} \quad (34)$$

Note that c_t and Γ both depend on ρ as can be seen from the definitions in (30) and (32). The use of an M-H step made it possible to handle (34) in a straightforward manner. A random walk proposal density was used. Note that if informative priors on ρ were used, for instance to capture the relationship between ρ and the cyclical order n, then the algorithm would remain the same, with the scale in the M-H step adjusted.

4. Analogously to the case for ρ, for the frequency parameter we have

$$p(\lambda_c|\sigma_\kappa^2, \sigma_\zeta^2, \sigma_\varepsilon^2, \rho, \alpha, Y) \propto p(\lambda_c)p(Y|\alpha, \theta)p(\alpha|\theta) \propto p(\lambda_c)p(\alpha|\theta) \quad (35)$$

$$p(\lambda_c|\sigma_\kappa^2, \sigma_\zeta^2, \sigma_\varepsilon^2, \rho, \alpha, Y) \propto p(\lambda_c)\left|\Gamma\right|^{-1/2} \exp\left\{-\frac{1}{2}\psi_1{}'\Gamma^{-1}\psi_1 - \frac{1}{2\sigma_\kappa^2}\sum_{t=2}^{T} c_t\right\} \quad (36)$$

With a beta density for $p(\lambda_c)$, the conditional posterior is nonstandard as trigonometric functions of λ_c occur within the exponent. Again, an M-H step was applied, with candidates generated by a random walk.

While 1. and 2. are sampled using a standard algorithm and direct Gibbs sampling, respectively, steps 3. and 4. require additional effort. The shape of these conditional posteriors seems unclear before estimation, and there is

little information to guide the choice of a problem-specific proposal density (this would make importance sampling more difficult). Chib and Greenberg (1996) discuss the application of the M-H sampler in econometric modelling and show examples where M-H steps are embedded within MCMC routines with multiple blocks. Two recent examples are Koop and van Dijk (2000) and Bos, Mahieu, and van Dijk (2000). The method is appropriate for the conditional densities of ρ and λ_c since the value of the normalizing constant is unknown, and we can only evaluate a constant multiple of the density ordinate.

We implemented random walk chains, and the scales were calibrated for each model and prior based on the suggested rule of thumb noted in Chib and Greenberg (1996). The variances for the random walk innovations were set to attain acceptance probabilities of 30-40%. The simulations performed well within this range, though the exact choice of scales is not crucial. The M-H properties were adapted to different priors, cyclical orders, and datasets, by calibrating the scales for ρ and λ_c to reach the target range of acceptance. A separate routine was designed for the Ox program so that the M-H scales could be adapted automatically in each case. The calibration was based on Gibbs sampling with a small number of iterations and typically required only a minute or so on a 500 MHz Pentium III PC.

In practice, an automatic calibration routine was included in the program to set these two quantities. Given preliminary values, say both equal to 0.1, the scales for the ρ and λ_c M-H steps were then progressively improved until the target range of acceptance was reached. Thus the M-H properties could be automatically adapted to different priors, cyclical orders, and datasets. The calibration was based on computing acceptance rates for Gibbs sampling with a small number of iterations, and typically required around a minute or so on a PC with a 500 MHz Intel Pentium III processor, depending on what scales were tried at the outset.

The ultimate set of drawings $\{\theta^{(j)}, \alpha^{(j)}\}$, $j = 1, ..., J$, used in the final posterior sample were a subset of the $\{\theta^{(i)}, \alpha^{(i)}\}$ produced by the MCMC routine. To improve the efficiency of the posterior samples, multiple iterations per posterior draw were run, and a certain number of initial iterations were discarded as the chain moved toward convergence. Each posterior sample ultimately consisted of 5,000 parameter draws, and apart from the slope variance, the correlations between successive draws in the final sample typically fell to near zero after just a few lags. The correlograms for the series of parameter draws is shown in figure 24 for real GDP with $n = 1$ using the

least informative prior. The correlogram for the period typically appeared similar to that of the frequency. The starting values for the parameters had a negligible impact on the final results assuming that plausible values were used. The correlations for higher order cyclical models and for the bivariate models and annual data showed similar decaying patterns with lag length.

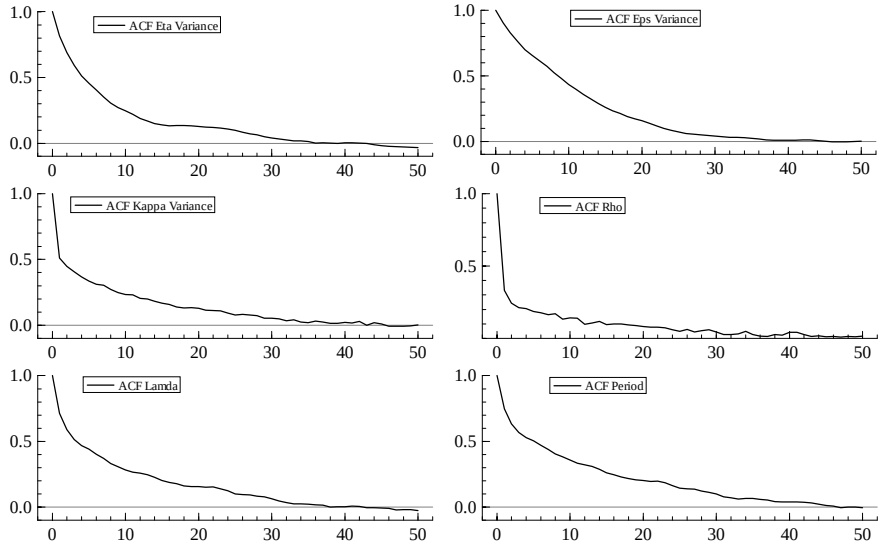

Figure 24: Correlations between successive posterior draws for US real GDP (logarithms) for $n = 1$ with least informative prior on λ_c.

Marginal likelihoods may be estimated using the posterior output. Kass and Raftery (1995) and DiCiccio, Kass, Raftery, and Wasserman (1997) discuss how to compute them using approaches based on the Laplace method. The Laplace method, which relies on a multivariate Gaussian approximation, is appealing due to its simplicity. For our particular MCMC study, exact analysis of complete conditional densities, carefully calibrated M-H steps, and multiple iterations per draw helped to produce a posterior sample with favourable properties. If it seems too costly or impractical to mitigate the influence of correlations and outliers in the draws, or if higher precision is desired, then the basic Laplace approximation can be adjusted in a variety of ways. For example, if the chain shows a tendency to veer away for long intervals, then one may use an estimator of the posterior covariance matrix that is robust to multivariate outliers.

Another sort of method is that of Chib (1995) and Chib and Jeliazkov (2001). In the earlier version, it is assumed that the complete conditional densities can be evaluated, which means the normalising constants are known. This is not the case for ρ and λ_c in our model. A simple idea might be to numerically integrate to find the normalising constants at each iteration, but this could be computationally inefficient. The framework in Chib and Jeliazkov (2001) applies to Gibbs samplers with M-H steps. As with Chib (1995), additional simulations must be run to estimate the quantities needed to compute the posterior ordinate, but now the properties of the M-H algorithm, in particular the reversibility of the transition kernel, are used to design the reduced chains.

Both methods require the evaluation of the prior and likelihood, preferably at a point of high density. We used the posterior mean, which is simple to compute. Alternatively, the posterior mode or median could be estimated. For the basic Laplace method, the estimated posterior ordinate is then obtained directly from the sample posterior covariance matrix. This approximation, which relies on a multivariate Gaussian kernel approximation to the posterior, is appropriate for comparing marginal likelihoods in cases where the posterior is relatively well-behaved, for instance it is not multimodal in a high density region for any of the parameters.

State matrix and Bayesian smoother

Taken on their own, the draws $\{\alpha^{(j)}\}$ are variates from $p(\alpha|Y)$. The posterior density of α involves a substantial amount of information as it describes the relationship among subsets of the state vector at various points in time. The associated marginal densities of the components are of primary interest for current purposes, but one could in theory examine virtually any aspect of the joint density of the cyclical estimates by averaging over different functions of draws. This permits a great deal of flexibility in analyzing the cycle and its relative positions at different points in time. The conditional mean of the cyclical component, given the data, represents the Bayesian counterpart of the classical smoother.

D Multivariate models

In this appendix we present a Bayesian treatment of the general similar cycle model, (12).

D.1 SSF and initial conditions

Interest centres on a Bayesian treatment of the cyclical component. The state vector for the nth order stochastic cycle in the multivariate model is

$$\begin{aligned}\psi_t &= [\psi^1_{n,t}, ..., \psi^N_{n,t}, \psi^{1*}_{n,t}, ..., \psi^{N*}_{n,t}, ...\psi^1_{1,t}, ..., \psi^N_{1,t}, \psi^{1*}_{1,t}, ..., \psi^{N*}_{1,t},]' \\ &= [\psi'_{n,t}, \psi'_{n-1,t}, ..., \psi'_{2,t}, \psi'_{1,t}]'\end{aligned}$$

With similar cycles in the N series, the transition equation is

$$\psi_t = (T_\psi \otimes I_N)\psi_{t-1} + c_n \otimes \begin{bmatrix} \kappa_t \\ \kappa^*_t \end{bmatrix} \tag{37}$$

where $\kappa_t = [\kappa^1_t, ..., \kappa^N_t]'$, $\kappa^*_t = [\kappa^{1*}_t, ..., \kappa^{N*}_t]'$ are vectors of the cyclical disturbances with the same covariance matrix, $\mathbf{\Sigma}_\kappa$, and T_ψ, the transition matrix in the univariate case, is given in equation (25). The initial conditions are formed by noting that the unconditional distribution of $\boldsymbol{\psi}_t$ has mean zero and variance matrix given by

$$\mathbf{\Sigma}_\psi = V(\rho, \lambda_c, n) \otimes \mathbf{\Sigma}_\kappa \tag{38}$$

where $V(\rho, \lambda_c, n)$ is given in (32).

The analytical result in (38) is important as it helps avoid potential numerical problems in the computationally more demanding multivariate framework. Since the variance matrix in the multivariate case may be immediately obtained from the expression in the univariate setup, there is no need to invert very high-dimensional matrices that may be near singular.

The proof for (38) is as follows. The cyclical error for the univariate transition equation (24) has the form $\sigma^2_\kappa B$ for the constant matrix $B = c_n c'_n \otimes I_2$. More generally, for N similar cycles, the covariance matrix of the error vector in (37) is $B \otimes \mathbf{\Sigma}_\kappa$.

The expression for the variance matrix of the cyclical state in the univariate case is $\sigma^2_\kappa V(\rho, \lambda_c, n)$, where $V(\rho, \lambda_c, n)$ satisfies the equation:

$$V(\rho, \lambda_c, n) = TV(\rho, \lambda_c, n)T' + B \tag{39}$$

Similarly, the cycle variance matrix in the multivariate case, $\mathbf{\Sigma}_\psi$, is the solution to:

$$\mathbf{\Sigma}_\psi = (T \otimes I_N)\mathbf{\Sigma}_\psi(T \otimes I_N) + B \otimes \mathbf{\Sigma}_\psi \tag{40}$$

The equivalence with (38) is seen by first taking the Kronecker product of both sides of (39) with Σ_κ to give

$$V(\rho, \lambda_c, n) \otimes \Sigma_\kappa = (TV(\rho, \lambda_c, n)T') \otimes \Sigma_\kappa + B \otimes \Sigma_\kappa \tag{41}$$

Further algebraic manipulation then yields (40).

The density of the initial state is now

$$p(\alpha_1|\Sigma_\kappa, \rho, \lambda_c) \propto |\Sigma_\psi|^{-1/2} \exp(-\frac{1}{2}\boldsymbol{\psi}_1{}'\boldsymbol{\Sigma}_\psi^{-1}\boldsymbol{\psi})$$

The conditional density of the state matrix, given the parameters, is

$$\begin{aligned} p(\alpha|\Sigma_\zeta, \Sigma_\kappa, \rho, \lambda_c) \quad \propto \quad & |\Sigma_\psi|^{-1/2} |I_2 \otimes \Sigma_\kappa|^{-(T-1)/2} |\Sigma_\zeta|^{-(T-1)/2} \\ & \times \exp\left\{ -\frac{1}{2}\boldsymbol{\psi}_1{}'\boldsymbol{\Sigma}_\psi^{-1}\boldsymbol{\psi}_1 - \frac{1}{2}\sum_{t=2}^{T} \mathbf{c}_t'(I_2 \otimes \boldsymbol{\Sigma}_\kappa)^{-1}\mathbf{c}_t \right\} \\ & \times \exp\left\{ -\frac{1}{2}\sum_{t=2}^{T} (\boldsymbol{\mu}_{1,t} - \boldsymbol{\mu}_{1,t-1})'\Sigma_\zeta^{-1}(\boldsymbol{\mu}_{1,t} - \boldsymbol{\mu}_{1,t-1}) \right\} \end{aligned} \tag{42}$$

where $\mathbf{c}_t = \boldsymbol{\psi}_{1,t} - (T \otimes I_N)\boldsymbol{\psi}_{1,t-1}$.

The summation in the cyclical part may be written as

$$\begin{aligned} \sum_{t=2}^{T} \mathbf{c}_t'(I_2 \otimes \boldsymbol{\Sigma}_\kappa)^{-1}\mathbf{c}_t \quad &= \quad \sum_{t=2}^{T} tr\left[(I_2 \otimes \boldsymbol{\Sigma}_\kappa^{-1})\mathbf{c}_t\mathbf{c}_t'\right] \\ &= \quad tr(\boldsymbol{\Sigma}_\kappa^{-1}\mathbf{c}_{1,t}\mathbf{c}_{1,t}') + tr(\boldsymbol{\Sigma}_\kappa^{-1}\mathbf{c}_{2,t}\mathbf{c}_{2,t}') \end{aligned}$$

where $\mathbf{c}_t' = \begin{bmatrix}\mathbf{c}_{1,t}' & \mathbf{c}_{2,t}'\end{bmatrix}$. For this partition of $\mathbf{c}_t$, the upper half $\mathbf{c}_{1,t}$ (N x N) corresponds to the N cycles, that is $\boldsymbol{\psi}_{1,t} = [\psi_{1,t}^1, ..., \psi_{1,t}^N]'$, and the lower half $\mathbf{c}_{2,t}$ to the auxiliaries, $\boldsymbol{\psi}_{1,t}^* = [\psi_{1,t}^{*1}, ..., \psi_{1,t}^{*N}]'$.

Thus the cyclical part of the conditional density of α may be written as

$$\begin{aligned} p(\alpha|\Sigma_\kappa, \rho, \lambda_c) \quad \propto \quad & |V(\rho, \lambda_c, n) \otimes \Sigma_\kappa|^{-1/2} |I_2 \otimes \Sigma_\kappa|^{-(T-1)/2} \\ & \times \exp\left\{ -\frac{1}{2}\boldsymbol{\psi}_1{}' \left[V(\rho, \lambda_c, n) \otimes \Sigma_\kappa\right]^{-1} \boldsymbol{\psi}_1 - \frac{1}{2} tr\left(\boldsymbol{\Sigma}_\kappa^{-1} G\right) \right\} \end{aligned}$$

where

$$G = \sum_{t=2}^{T} (\mathbf{c}_{1,t}\mathbf{c}'_{1,t} + \mathbf{c}_{2,t}\mathbf{c}'_{2,t})$$

Since

$$|V(\rho, \lambda_c, n) \otimes \Sigma_\kappa| = |V(\rho, \lambda_c, n)|^N |\Sigma_\kappa|^{2n}$$

and $|I_2 \otimes \Sigma_\kappa| = |\Sigma_\kappa|^2$ we have

$$|V(\rho, \lambda_c, n) \otimes \Sigma_\kappa|^{-1/2} |I_2 \otimes \Sigma_\kappa|^{-(T-1)/2} = |V(\rho, \lambda_c, n)|^{-N/2} |\Sigma_\kappa|^{-(n+T-1)}$$

Let the elements of block i, j of $V(\rho, \lambda_c, n)$ be denoted by

$$V(\rho, \lambda_c, n)_{i,j} = \begin{bmatrix} V(\rho, \lambda_c, n)_{i,j} & V(\rho, \lambda_c, n)_{i,j*} \\ V(\rho, \lambda_c, n)_{i*,j} & V(\rho, \lambda_c, n)_{i*,j*} \end{bmatrix}$$

The complete cyclical state vector at each t is partitioned as

$$\begin{aligned} \boldsymbol{\psi}_t &= [\psi^1_{n,t}, ..., \psi^N_{n,t}, \psi^{1*}_{n,t}, ..., \psi^{N*}_{n,t}, ...\psi^1_{1,t}, ..., \psi^N_{1,t}, \psi^{1*}_{1,t}, ..., \psi^{N*}_{1,t},]' \\ &= [\boldsymbol{\psi}'_{n,t}, \boldsymbol{\psi}^{*\prime}_{n,t}, \boldsymbol{\psi}'_{n-1,t}, ..., \boldsymbol{\psi}'_{1,t}, \boldsymbol{\psi}^{*\prime}_{1,t}]' \end{aligned}$$

The term involving the initial state vector in the exponent may be written as

$$\boldsymbol{\psi}_1{}' [V(\rho, \lambda_c, n) \otimes \Sigma_\kappa]^{-1} \boldsymbol{\psi}_1 = \frac{1}{2} tr(\Sigma_\kappa^{-1} H) \tag{43}$$

where the (N x N) matrix H is given by

$$\begin{aligned} H &= \sum_{i=1}^{n} (\sum_{j=1}^{n} V(\rho, \lambda_c, n)^{-1}_{i,j} \boldsymbol{\psi}_{j,1} \boldsymbol{\psi}'_{i,1} + \sum_{j=1}^{n} V(\rho, \lambda_c, n)^{-1}_{i,j*} \boldsymbol{\psi}^*_{j,1} \boldsymbol{\psi}'_{i,1}) \\ &\quad + \sum_{i=1}^{n} (\sum_{j=1}^{n} V(\rho, \lambda_c, n)^{-1}_{i*,j} \boldsymbol{\psi}_{j,1} \boldsymbol{\psi}^{*\prime}_{i,1} + \sum_{j=1}^{n} V(\rho, \lambda_c, n)^{-1}_{i*,j*} \boldsymbol{\psi}^*_{j,1} \boldsymbol{\psi}^{*\prime}_{i,1}) \end{aligned}$$

This expression is derived by first noting that

$$[V(\rho, \lambda_c, n) \otimes \Sigma_\kappa]^{-1} = tr\left[(V(\rho, \lambda_c, n)^{-1} \otimes \Sigma_\kappa^{-1}) \boldsymbol{\psi}_1 \boldsymbol{\psi}'_1\right]$$

Once the Kronecker product and multiplication are applied, the sum of the diagonal elements gives (43).

The summation for the trend part of the conditional density can similarly be written as

$$\sum_{t=2}^{T}(\boldsymbol{\mu}_{1,t}-\boldsymbol{\mu}_{1,t-1})'\Sigma_{\zeta}^{-1}(\boldsymbol{\mu}_{1,t}-\boldsymbol{\mu}_{1,t-1}) = \sum_{t=2}^{T} tr\left[\Sigma_{\zeta}^{-1}(\boldsymbol{\mu}_{1,t}-\boldsymbol{\mu}_{1,t-1})(\boldsymbol{\mu}_{1,t}-\boldsymbol{\mu}_{1,t-1})'\right]$$

In summary, the conditional density of the state matrix is

$$\begin{aligned} p(\alpha|\Sigma_{\kappa},\rho,\lambda_c) \quad \propto \quad & |V(\rho,\lambda_c,n)|^{-N/2}\,|\Sigma_{\kappa}|^{-(n+T-1)}\exp\left[-\frac{1}{2}tr(\Sigma_{\kappa}^{-1}[G+H])\right] \\ & \times|\Sigma_{\zeta}|^{-(T-1)/2}\exp\left\{-\frac{1}{2}tr\left(\Sigma_{\zeta}^{-1}\left[\sum_{t=2}^{T}(\boldsymbol{\mu}_{1,t}-\boldsymbol{\mu}_{1,t-1})(\boldsymbol{\mu}_{1,t}-\boldsymbol{\mu}_{1,t-1})'\right]\right)\right\} \end{aligned} \tag{44}$$

D.2 Priors

The priors on λ_c and ρ remain the same as for the univariate Bayesian treatment. Inverted Wishart priors are used for the variance matrices. This class of density represents a generalisation of the inverted gamma class, and the natural conjugate property extends to the multivariate model.

A random matrix Σ (N x N) is distributed as an inverted Wishart, denoted by $\boldsymbol{\Sigma} \sim IW(c,\mathbf{S})$, if for positive definite symmetric $\boldsymbol{\Sigma}$ the density function is given by

$$\begin{aligned} f(\boldsymbol{\Sigma}) &= C_{IW}(c,\mathbf{S};N)^{-1}\,|\boldsymbol{\Sigma}|^{-\frac{1}{2}(c+N+1)}\exp\left[-\frac{1}{2}tr(\boldsymbol{\Sigma}^{-1}\mathbf{S})\right], \\ C_{IW}(c,\mathbf{S};N) &= 2^{\frac{1}{2}cN}\pi^{\frac{1}{4}N(N-1)}\prod_{i=1}^{N}\Gamma(\frac{c-i+1}{2})\,|\mathbf{S}|^{-\frac{c}{2}} \end{aligned} \tag{45}$$

Note the similarity with the notation for the inverted gamma; in (45) $\mathbf{S}$ represents the shape matrix, and when the dimension $N = 1$, the formula reduces to (17).

Independent IW priors are assumed for the variance matrices with $c_{\zeta} = c_{\kappa} = c_{\varepsilon} = 10^{-7}$ and $\mathbf{S}_{\zeta} = \mathbf{S}_{\kappa} = \mathbf{S}_{\varepsilon} = 10^{-7}I_N$.

D.3 MCMC routine

The conditional density of α is shown in (44). The joint density of the observations given the hyperparameters and α is a direct extension of the univariate expression:

$$p(Y|\alpha,\theta) \propto |\Sigma_\varepsilon|^{-T/2} \exp\left\{ -\frac{1}{2}\sum_{t=1}^{T} tr\left[\Sigma_\varepsilon^{-1}(y_t - z_t\alpha_t)(y_t - z_t\alpha_t)'\right] \right\}$$

The joint posterior of $\{\Sigma_\zeta, \Sigma_\kappa, \Sigma_\varepsilon\}$ factors into a product of inverted Wishart densities with shape and scale matrix parameters given by

$$\begin{aligned}
\Sigma_\zeta &: \quad c_\zeta^* = c_\zeta + (T-1), \quad \mathbf{S}_\zeta^* = \mathbf{S}_\zeta + \sum_{t=2}^{T} (\boldsymbol{\mu}_{1,t} - \boldsymbol{\mu}_{1,t-1})(\boldsymbol{\mu}_{1,t} - \boldsymbol{\mu}_{1,t-1})' \\
\Sigma_\kappa &: \quad c_\kappa^* = c_\kappa + 2(n+T-1), \quad \mathbf{S}_\kappa^* = \mathbf{S}_\kappa + G + H \\
\Sigma_\varepsilon &: \quad c_\varepsilon^* = c_\varepsilon + T, \quad \mathbf{S}_\varepsilon^* = \mathbf{S}_\varepsilon + \sum_{t=1}^{T} (y_t - z_t\alpha_t)(y_t - z_t\alpha_t)'
\end{aligned}$$