Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 140 (2007) 618–649
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Trends and cycles in economic time series: A Bayesian approach

Andrew C. Harvey[a,*], Thomas M. Trimbur[b], Herman K. Van Dijk[c]

[a]*Faculty of Economics, Cambridge University, Sidgwick Avenue, Cambridge CB3 9DD, UK*
[b]*U.S. Bureau of the Census, Washington DC, UK*
[c]*Econometric Institute, Erasmus University, The Netherlands*

Available online 28 August 2006

Abstract

Trends and cyclical components in economic time series are modeled in a Bayesian framework. This enables prior notions about the duration of cycles to be used, while the generalized class of stochastic cycles employed allows the possibility of relatively smooth cycles being extracted. The posterior distributions of such underlying cycles can be very informative for policy makers, particularly with regard to the size and direction of the output gap and potential turning points. From the technical point of view a contribution is made in investigating the most appropriate prior distributions for the parameters in the cyclical components and in developing Markov chain Monte Carlo methods for both univariate and multivariate models. Applications to US macroeconomic series are presented.

JEL classification: C11; C32; E32

Keywords: Output gap; Kalman filter; Markov chain Monte Carlo; Real-time estimation; Turning points; Unobserved components

1. Introduction

Decomposing time series into trends and cycles is fundamental to a good deal of macroeconomic analysis. Key features of the 'business' cycle, such as length and turning points, are of great interest to policy makers in industry and government. Similarly,

*Corresponding author. Tel.: +44 1223 335 228; fax: +44 1223 335 475.
E-mail addresses: ach34@econ.cam.ac.uk (A.C. Harvey), hkvandijk@few.eur.nl (H.K. Van Dijk).

0304-4076/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2006.07.006

potential output trends and their deviations from the actual level of output yield important signals on the performance of an economy.

Trends and cycles may be modeled[1] directly as unobserved components within the framework of structural time series models (STMs); see Harvey and Jaeger (1993). The statistical treatment is based on the state space form with the components in a linear model extracted by the Kalman filter and associated smoother. For a Gaussian model, the likelihood function is obtained from the innovations produced by the Kalman filter and maximized numerically with respect to the unknown parameters. This classical procedure is implemented in the STAMP package of Koopman et al. (2000). However, fitting the standard trend plus cycle plus irregular model to time series of gross domestic product (GDP) often results in the irregular component disappearing with the result that the cycle is quite noisy. More generally, maximum likelihood (ML) can sometimes produce implausible parameter values, resulting perhaps in trends that are too inflexible or cycles that have too long a period. The higher order stochastic cycles introduced recently by Harvey and Trimbur (2003) tend to produce smoother extracted cycles, but problems of implausible estimates still remain. This provides one of the motivations for investigating a Bayesian approach.

A key parameter in the stochastic cycle is the period around which most of the power of the spectrum is concentrated. In building models to capture business cycles, it is not unreasonable to take on board prior notions about the period. These may be incorporated into the model in a flexible way; we do this here using a beta prior distribution. Previous work on using Bayesian methods for STMs, such as Durbin and Koopman (2002) and Koop and Van Dijk (2000), has not dealt with cycles. Huerta and West (1999a, b) study cyclical behavior indirectly using autoregressive models, but we would argue that our formulation is a more natural one for taking on board prior notions of periodicity and smoothness.

We present a Markov chain Monte Carlo (MCMC) algorithm in order to compute posterior results on parameters, model probabilities and unobserved components. The treatment of cycles introduces a number of new issues that need to be addressed. With the aid of state space modeling techniques, we set out an efficient procedure for computing the joint posterior density of parameters and components based on Gibbs sampling. In doing so we draw on earlier work on the efficient smoothing of unobserved components by, amongst others, Carter and Kohn (1994), Frühwirth-Schnatter (1994), De Jong and Shephard (1995) and Durbin and Koopman (2002).

While one of the potential advantages of a Bayesian approach is that it is able to avoid fitting implausible models, another is that it can yield more informative results. For example, many of the parameters in STMs are variances. The small sample behavior of ML estimators of such parameters is not easy to pin down, but it is certainly the case that distributions can be very far from normality. When the true value of a variance is zero, the asymptotic theory is nonstandard. One response to gauging

[1]Our concern here is with a model-based approach. Although detrending methods such as the Hodrick–Prescott filter are popular in macroeconomics, they can be misleading when used inappropriately, as argued in Harvey and Jaeger (1993) and Cogley and Nason (1995). The same is true of the more recent band-pass filter of Baxter and King (1999); see Harvey and Trimbur (2003) and Murray (2003). Furthermore, model-based approaches have the advantage that measures of uncertainty can be attached to the output.

the significance[2] of estimated variances is to use the bootstrap as documented in Stoffer and Wall (2004). The Bayesian approach provides another, possibly more attractive, line of attack by offering the opportunity to examine posterior distributions.

Another motive for investigating a Bayesian approach is that it allows for parameter uncertainty in the posterior distributions of components. This is important in the present context since one of our concerns is to present information on the size of the output gap as represented by the cycle. It is also straightforward to compute statistics such as the probability that the cycle, or its rate of change, is negative. Of course, a Bayesian treatment also allows for parameter uncertainty in the predictive distributions of future observations.

Applications are based on quarterly US macroeconomic time series from 1947 onwards. Univariate methods are illustrated using real GDP and features of the cycle such as its duration, turning points and time-varying amplitude are analyzed. We also demonstrate on-line analysis of the size and direction of the output gap. A multivariate model is then fitted to consumption, investment and GDP, so providing a contrast with the well-known study of King et al. (1991). Finally, a bivariate model of inflation and output, developed from the ideas of Kuttner (1994), is used to estimate the output gap by exploiting the Phillips curve relationship.

The paper is arranged as follows. Section 2 reviews the extension of the class of STMs to include higher order cyclical components and discusses ways of capturing their features and assessing the size and direction of the output gap. The Bayesian treatment is developed in Section 3, while Section 4 applies the methods to US GDP. The multivariate applications are in Section 4.4 and Section 5 concludes. Technical details on the state space form and MCMC algorithms are laid out in an Appendix.

2. STM for trends and cycles

Define the $N \times 1$ vector of observations $\mathbf{y}_t$, where $\mathbf{y}_t = (y_t^1, \ldots, y_t^N)'$. The class of multivariate STMs under consideration consists of trend, cycle and irregular components, denoted by $N \times 1$ vectors $\boldsymbol{\mu}_t \boldsymbol{\psi}_{n,t}$ and $\boldsymbol{\varepsilon}_t$, respectively. Thus

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\psi}_{n,t} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim NID(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon), \quad t = 1, \ldots, T, \tag{1}$$

where $NID(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$ denotes that the vector is serially independent and normally distributed with zero mean vector and $N \times N$ positive semi-definite covariance matrix, $\boldsymbol{\Sigma}_\varepsilon$. The stochastic trend is a multivariate integrated random walk

$$\begin{aligned} \boldsymbol{\mu}_t &= \boldsymbol{\mu}_{t-1} + \boldsymbol{\beta}_{t-1}, \\ \boldsymbol{\beta}_t &= \boldsymbol{\beta}_{t-1} + \boldsymbol{\zeta}_t, \quad \boldsymbol{\zeta}_t \sim NID(0, \boldsymbol{\Sigma}_\zeta), \end{aligned} \tag{2}$$

where $\boldsymbol{\beta}_t$ is the vector of slopes. A seasonal component can easily be added if appropriate.

The vector $\boldsymbol{\psi}_{n,t}$ is a generalization of the similar cycle model of Harvey and Koopman (1997). The aim of the generalization, originally proposed by Harvey and Trimbur (2003), is to include higher order models that tend to produce smoother extracted cycles.

[2]Valid tests of the null hypothesis that the variance is zero can be carried out, but this usually requires that the restricted model be estimated; see Harvey (2001).

2.1. Generalized stochastic cycles

An nth-order univariate cycle is defined by

$$\begin{bmatrix}\psi_{1,t}\\ \psi^*_{1,t}\end{bmatrix}=\rho\begin{bmatrix}\cos\lambda_c & \sin\lambda_c\\ -\sin\lambda_c & \cos\lambda_c\end{bmatrix}\begin{bmatrix}\psi_{1,t-1}\\ \psi^*_{1,t-1}\end{bmatrix}+\begin{bmatrix}\kappa_t\\ \kappa^*_t\end{bmatrix},$$
$$\begin{bmatrix}\kappa_t\\ \kappa^*_t\end{bmatrix}\sim NID\left(\begin{bmatrix}0\\ 0\end{bmatrix},\begin{bmatrix}\sigma^2_\kappa & 0\\ 0 & \sigma^2_\kappa\end{bmatrix}\right), \tag{3}$$

$$\begin{bmatrix}\psi_{i,t}\\ \psi^*_{i,t}\end{bmatrix}=\rho\begin{bmatrix}\cos\lambda_c & \sin\lambda_c\\ -\sin\lambda_c & \cos\lambda_c\end{bmatrix}\begin{bmatrix}\psi_{i,t-1}\\ \psi^*_{i,t-1}\end{bmatrix}+\begin{bmatrix}\psi_{i-1,t-1}\\ \psi^*_{i-1,t-1}\end{bmatrix},\quad i=2,\ldots,n. \tag{4}$$

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, κ_t and κ^*_t, are assumed to be uncorrelated with each other and with the disturbances driving the other components. Harvey and Trimbur (2003) show that as n increases the signal extraction filter for a cycle plus noise model tends toward a band-pass filter as in Baxter and King (1999). General expressions for the variance, autocovariances and spectrum are given in Trimbur (2006).

In a similar cycle model ρ and λ_c are the same across all series. Therefore, the cycles have the same dynamic properties in the sense that their autocorrelation functions and spectral densities are identical. However, the cycles themselves are not, in general, identical. The similar cycle model, originally formulated for $n=1$, may be extended to higher order cycles by defining a $2nN\times 1$ state vector

$$\boldsymbol{\psi}_t=[\psi^1_{n,t},\ldots,\psi^N_{n,t},\psi^{1*}_{n,t},\ldots,\psi^{N*}_{n,t},\ldots\psi^1_{1,t},\ldots,\psi^N_{1,t},\psi^{1*}_{1,t},\ldots,\psi^{N*}_{1,t}]', \tag{5}$$

where the sub-vector $\boldsymbol{\psi}_{n,t}=[\psi^1_{n,t},\ldots,\psi^N_{n,t}]'$ appears in (1). Define the matrix

$$\mathbf{T}_n=\mathbf{I}_n\otimes\mathbf{T}+\mathbf{S}_n\otimes\mathbf{I}_2, \tag{6}$$

where

$$\mathbf{T}=\rho\begin{bmatrix}\cos\lambda_c & \sin\lambda_c\\ -\sin\lambda_c & \cos\lambda_c\end{bmatrix} \tag{7}$$

and $\mathbf{S}_n$ is $n\times n$ with ones on the off-diagonal strip that lies adjacent to the main diagonal on the right-hand side and zeros everywhere else; that is, the row i, column $i+1$ element of $\mathbf{S}_n$ equals 1 for $i=1,\ldots,n-1$, and all other elements equal 0. Define $\mathbf{c}_n$ to be an $n\times 1$ vector with one in the last position and zeros elsewhere. Then

$$\boldsymbol{\psi}_t=(\mathbf{T}_n\otimes\mathbf{I}_N)\boldsymbol{\psi}_{t-1}+\mathbf{c}_n\otimes\begin{bmatrix}\boldsymbol{\kappa}_t\\ \boldsymbol{\kappa}^*_t\end{bmatrix}, \tag{8}$$

where the assumptions on the $N\times 1$ vectors of Gaussian disturbance, $\boldsymbol{\kappa}_t$ and $\boldsymbol{\kappa}^*_t$, are

$$\mathrm{E}(\boldsymbol{\kappa}_t)=\mathbf{0},\quad \mathrm{E}(\boldsymbol{\kappa}_t\boldsymbol{\kappa}_t')=\mathrm{E}(\boldsymbol{\kappa}^*_t\boldsymbol{\kappa}^{*'}_t)=\boldsymbol{\Sigma}_\kappa,\quad \mathrm{E}(\boldsymbol{\kappa}_s\boldsymbol{\kappa}_t')=\mathbf{O}\ \text{ for } s\neq t$$

with $\boldsymbol{\Sigma}_\kappa$ an $N\times N$ covariance matrix and $\mathrm{E}(\boldsymbol{\kappa}_s\boldsymbol{\kappa}^{*'}_t)=\mathbf{O}$ for all $s,t=1,\ldots,T$.

2.2. The changing output gap and turning points in the cycle

The trend in GDP is often regarded as a permanent component, while the cycle is transitory, but serially correlated; see, for example, Blanchard and Fischer (1989). The STM makes the concept operational in that the trend is the component that yields the long-run forecasts; see Harvey (1989, pp. 284–286). This avoids the ad hoc nature of a trend based on the Hodrick–Prescott or band-pass filter. In what follows we will associate the trend with potential output and the cycle with the output gap.

Having fitted a model, we are interested in studying the characteristics of the cycle from the smoothed estimates. For example, although the expected value of the square of the amplitude is $E(\psi_t^2 + \psi_t^{*2}) = 2\sigma_\psi^2$, it may be of interest to plot estimates of the amplitude through time to see if it is changing in any way. Of course, the full distribution of the output gap at any time may be of prime interest and we may wish to use this to produce statistics such as the probability that the economy is below potential output.

One characteristic of a cyclical series is its turning points. (For instance, the change from top to bottom in a downturn of the cycle gives a measure of the severity of the contraction, but to measure this we clearly need to know the top and bottom). Here turning points are identified from the extracted cycle rather than from the application of a criterion directly to the series itself. A fairly straightforward approach is to follow Zellner et al. (1990) in labeling a time point t as a peak, or downturn, if

$$\widehat{\psi}_{n,t-b}, \widehat{\psi}_{n,t-b+1}, \ldots, \widehat{\psi}_{n,t-2}, \widehat{\psi}_{n,t-1} < \widehat{\psi}_{n,t} > \widehat{\psi}_{n,t+1}, \widehat{\psi}_{n,t+2}, \ldots, \widehat{\psi}_{n,t+a}, \tag{9}$$

where a and b are positive integers, and defining a trough, or upturn, in an analogous fashion. The smoother is the extracted cycle, $\widehat{\psi}_{n,t}$, the easier it should be to identify meaningful turning points.

The NBER turning points are defined with respect to the level of the series. Harding and Pagan (2002) show that a rather simple dating rule applied to the differences in GDP reproduces the NBER peaks and troughs quite closely. A trough at time t is defined by $\{\Delta_2 y_t < 0,\ \Delta y_t < 0,\ \Delta y_{t+1} > 0,\ \Delta_2 y_{t+2} > 0\}$ where $\Delta_2 y_{t+2} = y_{t+2} - y_t = \Delta y_{t+2} + \Delta y_{t+1}$, and a peak similarly. Expansions and contractions are defined from these peaks and troughs and used as the basis for recognizing a cycle and measuring its characteristics. The same criterion could be applied to the extracted level and cycle components by looking at changes in the estimates of $\mu_t + \psi_{n,t}$.

Information on the rate of change of the cycle is contained in the estimated state sub-vector (5). It is shown in Appendix A that the discrete time analogue of the expected incremental change of a continuous time first-order cycle is given by

$$D\psi_t = (\log \rho)\psi_t + \lambda_c \psi_t^*. \tag{10}$$

The second-order case is more complicated but it is argued in Appendix A that for the discrete time model formulated in the previous sub-section an appropriate measure of change is

$$D\psi_{2,t} = (\log \rho)\psi_{2,t} + \lambda_c \psi_{2,t}^* + \rho^{-1}(\psi_{1,t} \cos \lambda_c - \psi_{1,t}^* \sin \lambda_c). \tag{11}$$

This formula can also be used for higher order cycles with the subscripts 2 and 1 replaced by n and $n-1$, respectively. Plotting an estimate of $D\psi_{2,t}$ should be highly informative. So too should the series on the probability that $D\psi_{2,t}$ is negative (positive), this being the probability that the cycle is moving down (up).

Turning points could be associated with points at which the estimate of $D\psi_{2,t}$ changes sign. However, it transpires that $D\psi_{2,t}$ is too variable for this to be useful. One might, of course, question the whole notion of the identification of turning points and the associated binary classification into expansions and recessions. After all, the series of estimates of $\psi_{2,t}$ and $D\psi_{2,t}$ contain more information as they measure the size and direction of the output gap at all points in time. Even more information is contained in the full posterior distributions.

On-line tracking of components is of considerable importance for policy makers. With a model-based approach, information on the current cycle is obtained by filtering, but more accurate information is given by smoothing as new observations become available. The availability of posterior distributions allows the user to be fully aware of the uncertainty attached to any estimates; see the discussion in Orphanides and Van Norden (2002).

3. Bayesian treatment

For convenience, we present the univariate case in this section and refer to the Appendix for the multivariate case. The three variance parameters and two cyclical parameters are arranged in the vector $\boldsymbol{\theta} = \{\sigma_\zeta^2, \sigma_\kappa^2, \sigma_\varepsilon^2, \rho, \lambda_c\}$. The goal is to analyze the posterior distribution, $p(\boldsymbol{\theta}|\mathbf{y})$, where $\mathbf{y} = \{y_1, \ldots, y_T\}$ denotes the observations. The joint posterior of the trend and cycle components is produced as a by-product of the MCMC routine.

3.1. Priors

The direct interpretation of the cycle parameters in the STM makes it straightforward to design priors that reflect knowledge of the business cycle. Thus, for quarterly data we consider priors for frequency, λ_c, based on beta distributions with a mode at $2\pi/20$, corresponding to a period of five years. Fig. 1 shows three such priors, labeled wide, intermediate and sharp. For technical details see Appendix B.

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. Although it seems, on the basis of empirical work, that ρ falls as n increases, it is difficult to be precise about the form of the relationship so we use a uniform prior on ρ over the interval $[0, 1]$.

For the variance parameters independent flat priors on $[0, \infty]$ are assumed. The use of flat priors has the same advantage as inverted gammas of allowing one to sample directly, but in applications where expectations on the values of variance parameters are rather vague, a flat prior ensures that the likelihood surface is not distorted near zero.[3]

3.2. Posterior

The posterior distribution is obtained as

$$p(\boldsymbol{\theta}|\mathbf{y}) = L(\boldsymbol{\theta}; \mathbf{y})p(\boldsymbol{\theta}),$$

[3]The unobserved components models we consider in this paper have essentially the same fundamental structure as hierarchical models for studying group effects. As Gelman (2006) notes for the case of a basic hierarchical model, any noninformative prior on the group-level variance gives a proper posterior when sufficient data are available, as is the case here.

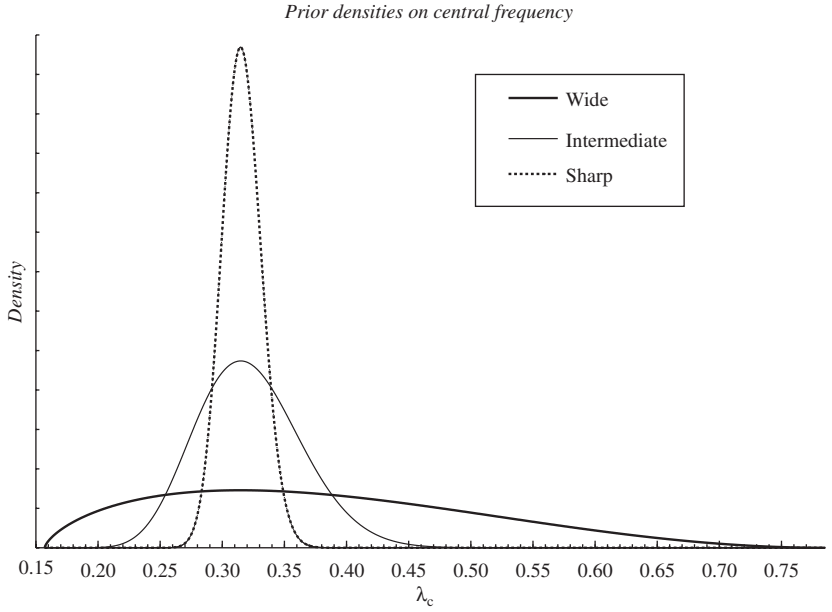

Fig. 1. Beta-based priors on λ_c, with mode equal to $2\pi/20$ (five-year period for quarterly data).

where the likelihood function, $L(\boldsymbol{\theta}; \mathbf{y})$, is evaluated using the Kalman filter; see, for example, Harvey (1989, p. 126). The posterior is difficult to work with directly as the constant of proportionality is not available analytically. A strategy is therefore needed for analyzing its properties. In the applications that follow we deal with up to 20-dimensional parameter vectors and MCMC methods offer an efficient way to sample (pseudo-random) parameter drawings from the posterior. This also allows us to produce drawings of regular functions of the parameters, such as periods of cycles and signal–noise ratios, and to compare finite sample results on posterior moments with ML estimates.

The parameter space is extended to include the components and associated auxiliary processes in (1), which together form the state vector, $\boldsymbol{\alpha}_t$, for the model, taken over all observation times $t = 1, \ldots, T$. The state space form is discussed in Appendix C. Our MCMC routine uses the simulation smoother for drawing from the conditional density of the state vector, as in De Jong and Shephard (1995), and, more recently, Durbin and Koopman (2002).

3.3. Signal extraction on characteristics of cyclical and trend components

The MCMC scheme produces draws from the joint posterior of the trend and cyclical components. The estimated component series are obtained by averaging over the J state draws, for example $\widehat{\mu}_t = \sum_{j=1}^{J} \mu_t^{(j)}/J$, where $\mu_t^{(j)}$ denotes the jth draw for the trend at time t. Higher order moments can be similarly constructed, while the amplitude of the cycle is given by

$$A_t = \frac{1}{J}\sum_{j=1}^{J} \sqrt{\widehat{\psi}_{n,t}^{2(j)} + \widehat{\psi}_{n,t}^{*2(j)}}, \quad t = 1, \ldots, T. \tag{12}$$

The probabilities that the cycle and its change, (11), are negative are straightforward to estimate.

In keeping with state space terminology we will refer to posterior means computed over the whole sample as smoothed estimates while the corresponding estimates based only on current and past observations will be called filtered estimates.

3.4. Model evaluation

The marginal likelihood is

$$M(\mathbf{y}) = \int L(\boldsymbol{\theta}; \mathbf{y}) p(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta}. \tag{13}$$

Bayes factors are computed as the ratio of marginal likelihoods for different model specifications $\{L(\boldsymbol{\theta}; \mathbf{y}), p(\boldsymbol{\theta})\}$. Posterior odds may be formed by multiplying Bayes factors by prior probability ratios that give the relative preference, ex ante, for the various likelihood-prior model structures; see Kass and Raftery (1995). The method we use for estimating marginal likelihoods is set out in Appendix C.

4. US GDP

In this section we fit univariate models, with different orders for the cycle component, to the logarithm of quarterly real GDP from 1947Q1 to 2004Q4 (source: Bureau of Economic Analysis, US Department of Commerce). The aim of the analysis is to present results on the mapping from prior to posterior of model parameters including a sensitivity analysis with respect to the choice of the prior on the frequency as shown in Fig. 1. The MCMC routine was written in the Ox language of Doornik (1999) with use being made of the state space algorithms in the SsfPack set of routines developed by Koopman et al. (1999). The properties of extracted cycles are analyzed by looking at turning points and changes in amplitude, as well as size and direction. Table 1 summarizes the results of fitting models in terms of posterior means and marginal likelihoods. Note that $2\pi/\lambda_c$ is the period (in quarters), σ_ψ^2 is the variance of the cycle and q, defined as $\sigma_\zeta^2/(\sigma_\psi^2 + \sigma_\varepsilon^2)$, is the ratio of the variance of the noise driving the trend to the variance of the stationary component. All variance parameters are multiplied by 10^7 while the estimated logarithm of the marginal likelihood is denoted by $m(y)$.

4.1. Priors and posteriors

Marginal posterior densities for $n = 1$, with the wide prior on λ_c, are shown in Figs. 2 and 3, along with 95% HPD (highest posterior density) regions. Each estimated density function from the MCMC routine represents a standard approximation based on local Gaussian kernels, as implemented in Doornik (1999). The 95% HPD region is defined as the interval of minimum length that contains 95% of the probability mass for which the upper and lower boundary have equal density. The HPD regions resemble classical confidence intervals, but have a different interpretation in that they represent a direct probability statement about the value of an uncertain parameter or hidden component.

Fig. 2 shows that the marginal posteriors of both the irregular variance, σ_ε^2, and the slope variance, σ_ζ^2, are skewed, but while the density of σ_ε^2 is concentrated near zero, that of σ_ζ^2

Table 1
Posterior means of parameters estimated for series on logarithms of quarterly US real GDP from 1947Q1 to 2004Q4 for cycles of order $n = 1$–4 with different priors on λ_c

n	Prior	σ_ζ^2	σ_κ^2	σ_ε^2	ρ	λ_c	$2\pi/\lambda_c$	σ_ψ^2	q	$m(y)$
1	Wide	46.1	466	32	0.884	0.409	16.02	2336	0.0276	698.1
	Intermediate	28.3	539	23	0.894	0.341	18.66	2911	0.0117	698.4
	Sharp	24.5	561	21	0.898	0.320	19.69	3110	0.0092	698.4
2	Wide	17.1	363	111	0.697	0.272	24.62	4603	0.0058	704.0
	Intermediate	20.2	322	117	0.695	0.308	20.71	3854	0.0068	704.1
	Sharp	20.5	314	118	0.694	0.313	20.09	3706	0.0070	704.0
3	Wide	26.5	218	148	0.560	0.291	23.29	4097	0.0115	703.5
	Intermediate	26.9	199	150	0.562	0.311	20.54	3723	0.0112	703.7
	Sharp	26.6	196	151	0.563	0.314	20.05	3648	0.0108	703.8
4	Wide	43.0	159	157	0.461	0.310	22.13	3804	0.0287	702.3
	Intermediate	37.8	151	159	0.467	0.314	20.36	3759	0.0202	702.8
	Sharp	37.5	150	159	0.468	0.314	20.04	3728	0.0206	702.9

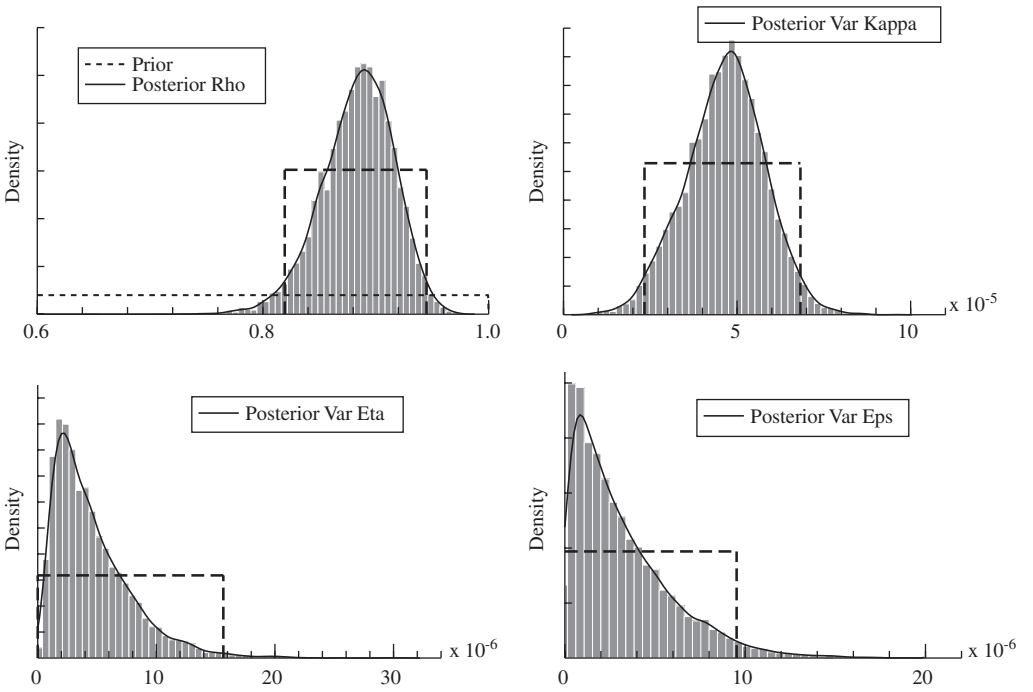

Fig. 2. Marginal posterior densities of ρ, σ_κ^2, σ_ζ^2, and σ_ε^2 for $n = 1$, with least informative prior on λ_c, for US GDP. The dashed lines indicate 95% highest posterior density (HPD) intervals.

displays a clear peak away from zero, giving clear evidence for stochastic variation in the trend. As regards the cycle parameters, the marginal posterior for ρ, which is based on a uniform prior on $[0, 1]$, peaks near 0.9, while the density of the cyclical error variance σ_κ^2 appears symmetric. The prior and posterior densities for the frequency, λ_c, and period,

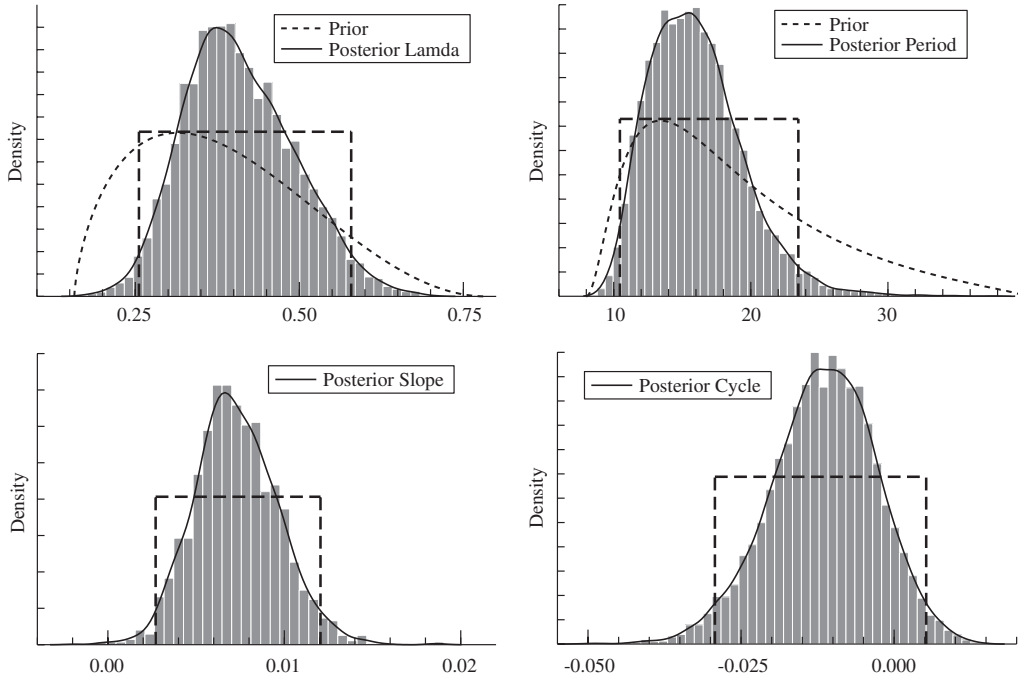

Fig. 3. (Top) Marginal posterior densities of λ_c and $2\pi/\lambda_c$ for $n = 1$ with wide informative prior on λ_c. (Bottom) Marginal posterior densities of the slope and cyclical components, $\beta_{t,T}$ and $\psi_{t,T}$, for 2003Q1.

$2\pi/\lambda_c$, appear in Fig. 3. The posteriors indicate a clear peak around a four- to five-year period even with a relatively noninformative prior. These results suggest that the likelihood surface in the first-order model has a more or less regular shape so that it is relatively straightforward to pick out a business cycle component.

The underlying growth rate (slope of the trend) and the cycle, estimated as described in Section 3.3, are displayed in Figs. 4 and 5. The shaded regions in Fig. 5 denote recessions as identified by the NBER. The marginal posterior densities of the slope and cycle at 2003Q1 are shown in the lower panels of Fig. 3 both appear to be more or less symmetric. The cycle will be analyzed in the next sub-section. Fig. 4 suggests the intriguing possibility of a cycle in the growth rate but this idea will not be pursued further in this paper.

For $n = 1$ a flat prior on λ_c gives nearly the same results as for the wide prior, since the likelihood surface with respect to λ_c has a clear peak. For higher order cycles, the results are more sensitive to the prior. With a flat prior the posterior of λ_c for $n = 2$ has most of its probability mass at very low values, with no clearly discernible peak.[4] A corresponding problem arises in computing ML estimates; Harvey and Trimbur (2003) had to resort to fixing λ_c in this case. Using an informative prior addresses the issue in a more flexible way. The results in Table 1 indicate that the differences between the outcomes for wide, intermediate and sharp priors are not great and so it is probably safest to stick to the wide prior.

[4]This figure is not shown here, but like others referred to in the text it was included in an Appendix to an earlier version of the paper and is available on request.

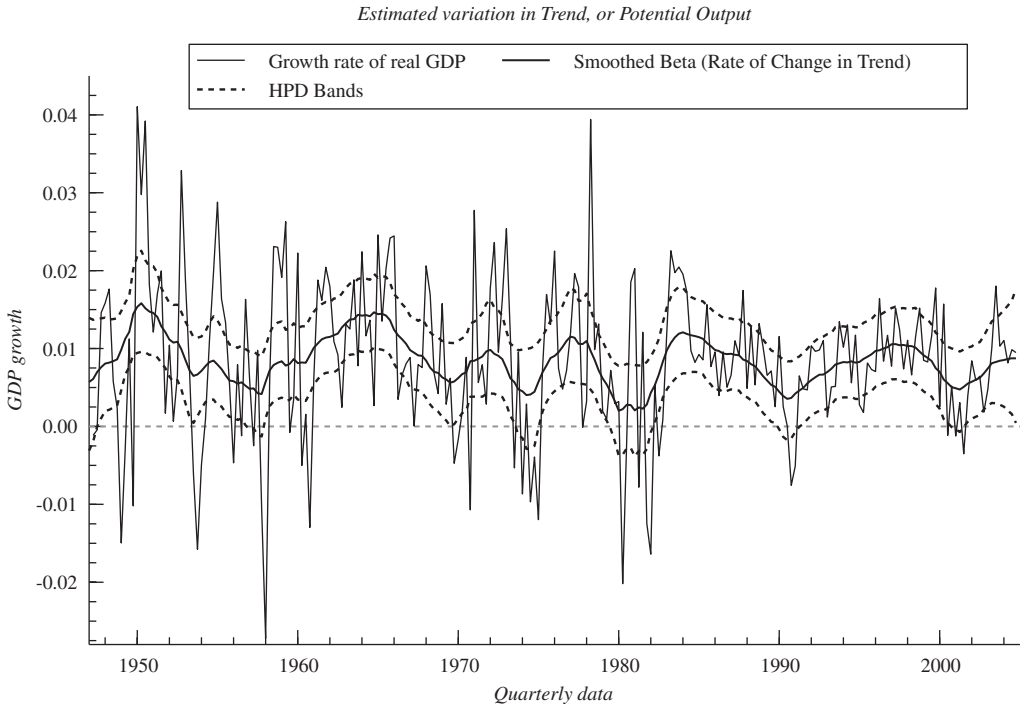

Fig. 4. Smoothed slope (growth rate) in trend of US GDP, with 95% HPD bands, estimated with $n = 1$ and a wide prior on λ_c.

Fig. 6 shows marginal posteriors of the cycle parameters for $n = 2$ with the wide prior. The posterior means are now around six years. The posteriors of σ_ε^2 and the cycle variance, σ_ψ^2, indicate[5] that more high-frequency movement is consigned to the irregular. Fig. 7 also shows the posterior distribution of the signal–noise ratio $q = \sigma_\zeta^2/(\sigma_\psi^2 + \sigma_\varepsilon^2)$. This measures the relative variation in the nonstationary and stationary parts of the model. The density in Fig. 7 is skewed with a 95% HPD region that extends up to 0.02.

Similar analysis could be carried out for $n = 3$ and 4. However, the main contrast is between $n = 2$ and the standard first-order case. Furthermore, the marginal likelihood shows the second-order cycle to be the preferred one.[6] The analysis in the next two subsections is based on the second-order cycle with a wide informative prior on the frequency.

4.2. Characteristics of the cycle

The relative smoothness of the second-order cycle shown in Fig. 8 makes it easier to identify peaks and troughs. We show the turning points evaluated from definition (9), with

[5]The posterior of σ_ψ^2 was obtained by computing the cycle variance for each set of draws $\{\rho^{(j)}, \lambda_c^{(j)}, \sigma_\kappa^2\}$.

[6]There is tricky theoretical issue here because of the use of diffuse priors on certain parameters. However, it is generally accepted that comparisons are valid so long as the parameters in question are not restricted and occur in all models; see Gelman (2006). In the present context our preference for $n = 2$ is supported by the plausibility of posterior moments, the properties of the extracted cycles and forecasting performance.

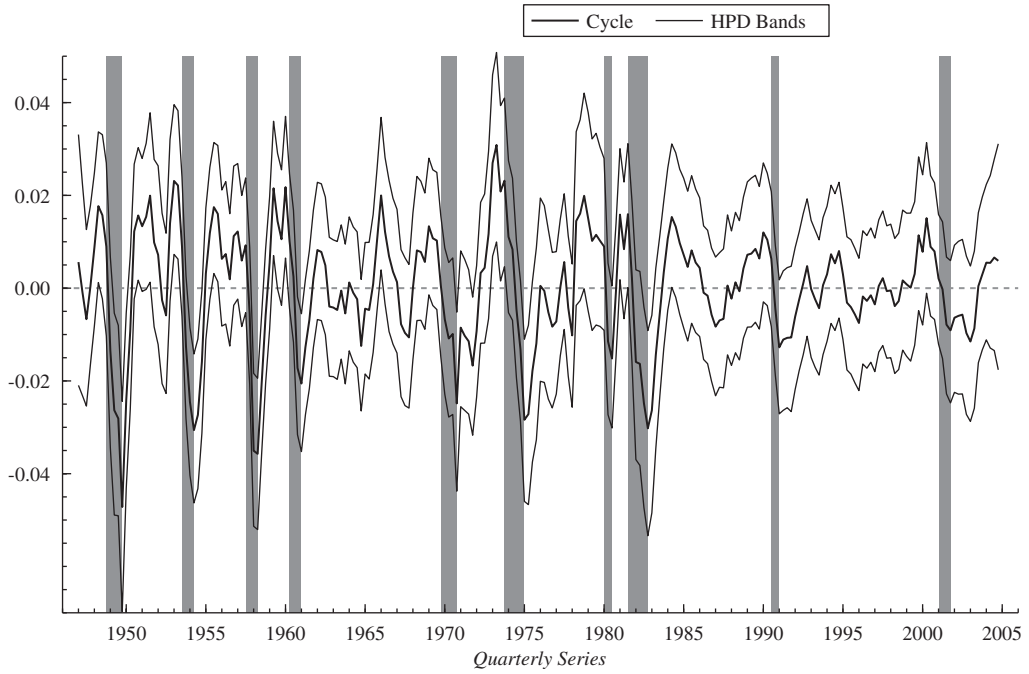

Fig. 5. First-order cycle in quarterly US real GDP, with 95% HPD bands, estimated using a wide informative prior on λ_c.

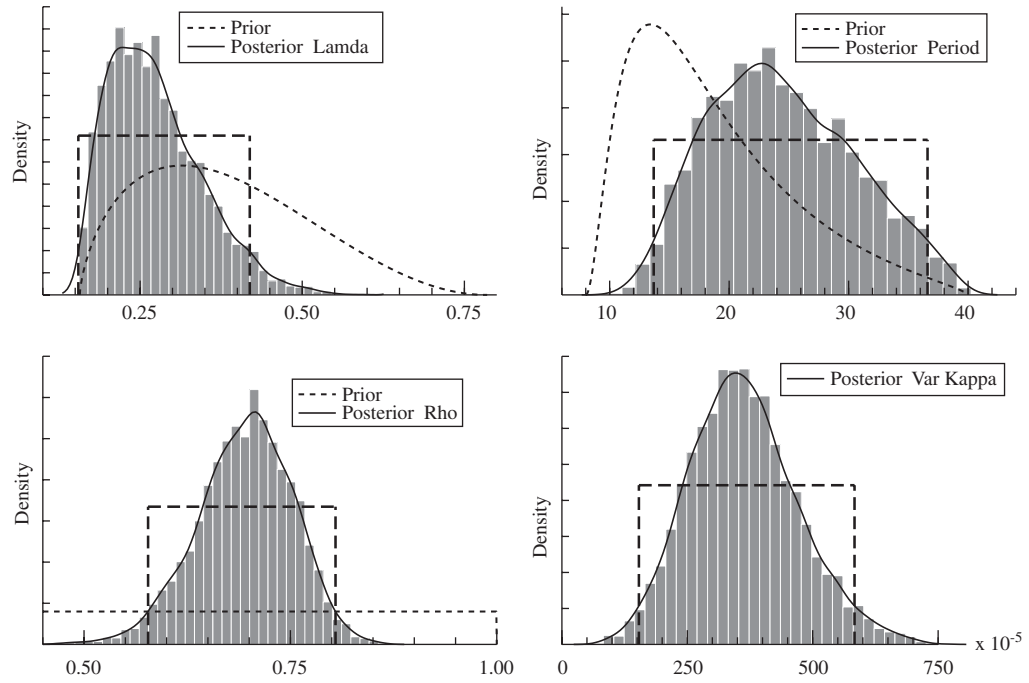

Fig. 6. Marginal posterior densities of US GDP cycle parameters for $n = 2$ with wide prior on λ_c.

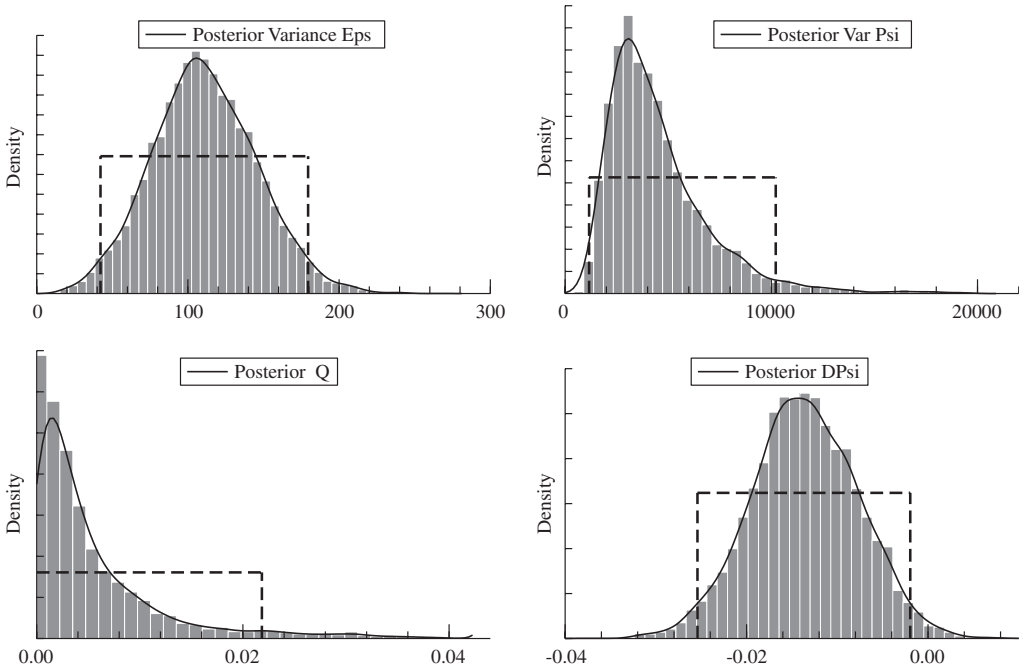

Fig. 7. (Top) Marginal posterior densities of irregular and cycle variances for $n = 2$ with wide prior on λ_c. (Bottom, left) Marginal posterior of signal–noise ratio, q. (Bottom, right) Marginal posterior of $D\psi_{2,t}$ for 1973Q4.

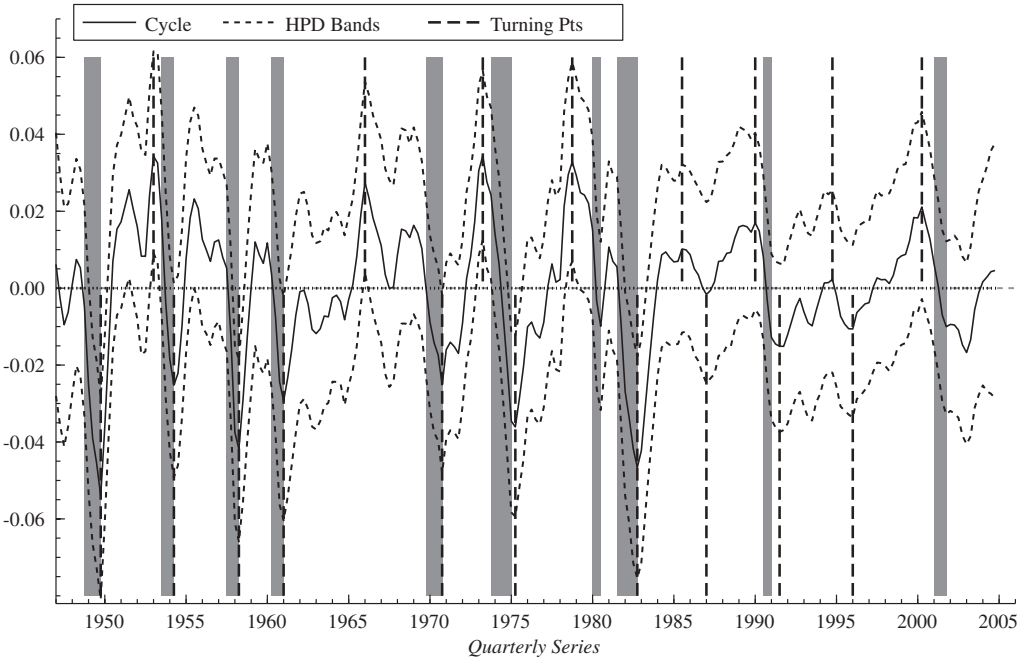

Fig. 8. Second-order cycle in quarterly US real GDP, with 95% HPD bands, estimated using a wide informative prior on λ_c.

$b = 10$ and $a = 8$. The peaks occur a little before the start of an NBER recession (shaded on the graph) while the troughs tend to match the end. However, as was observed in Section 2.2, focussing on the binary labeling of business cycle phases means that one loses sight of the finer detail offered by the extracted cycle. For example, the smoothed cycle in Fig. 8 shows clearly that some recessions are deeper than others. A plot of the changing amplitude of the cycle, using formula (12), allows one to see how the intensity of business cycles has changed over the last half-century. Fig. 9 displays the evolving amplitude. The strength of the cycle has been more moderate, as well as less volatile, since the mid-1980s. However, it shows a slight upturn in recent years, indicating that, contrary to the view of some economists, the business cycle is still relevant.

Fig. 10 shows a plot of the posterior mean of $D\psi_{2,t}$ together with a corresponding estimate based on $(\Delta\psi_t + \Delta\psi_{t+1})/2$. This second estimate yields a series very close to $D\psi_{2,t}$, though the averaging makes it slightly less volatile. The attraction of $D\psi_{2,t}$ is that it can be calculated at time t, rather than at $t+1$, and this is important at the end of the sample.

The bottom-right graph of Fig. 7 shows the marginal posterior of the change in the cycle, $D\psi_{2,t}$, for 1973Q4. As is clear from Fig. 8 this quarter marks the onset of the deep recession induced by the first oil price shock of the 1970s.

Table 2 shows the probability that $D\psi_{2,t}$ is negative over the four-year period surrounding this recession.

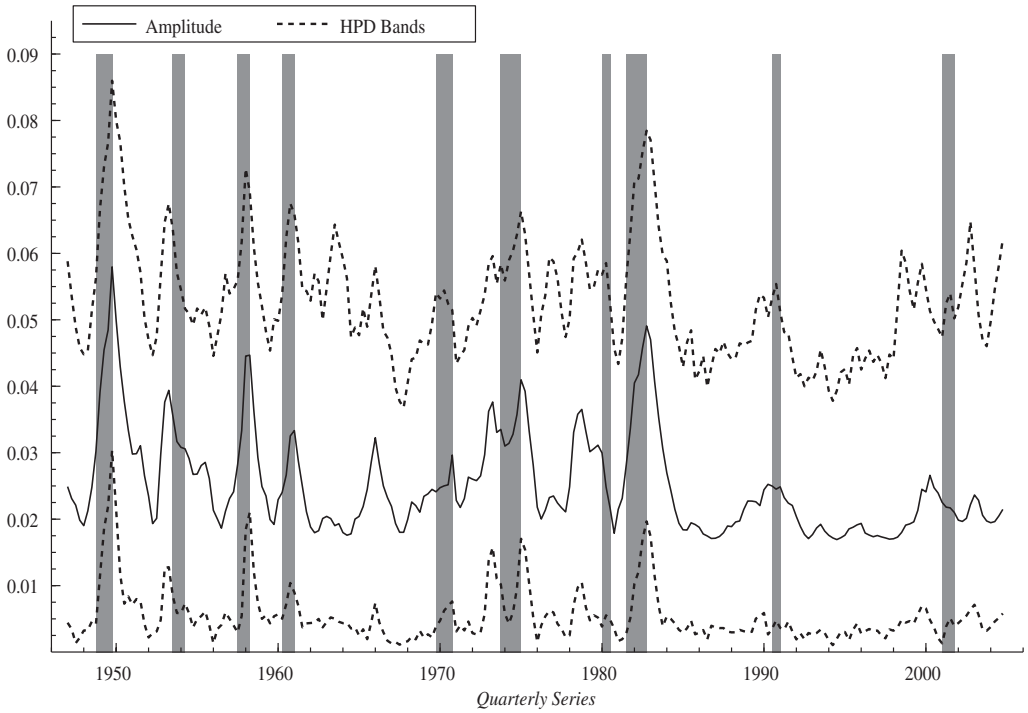

Fig. 9. Estimated amplitude of the cycle in quarterly US real GDP for $n = 2$ and a wide informative prior on λ_c, shown with 95% HPD bands.

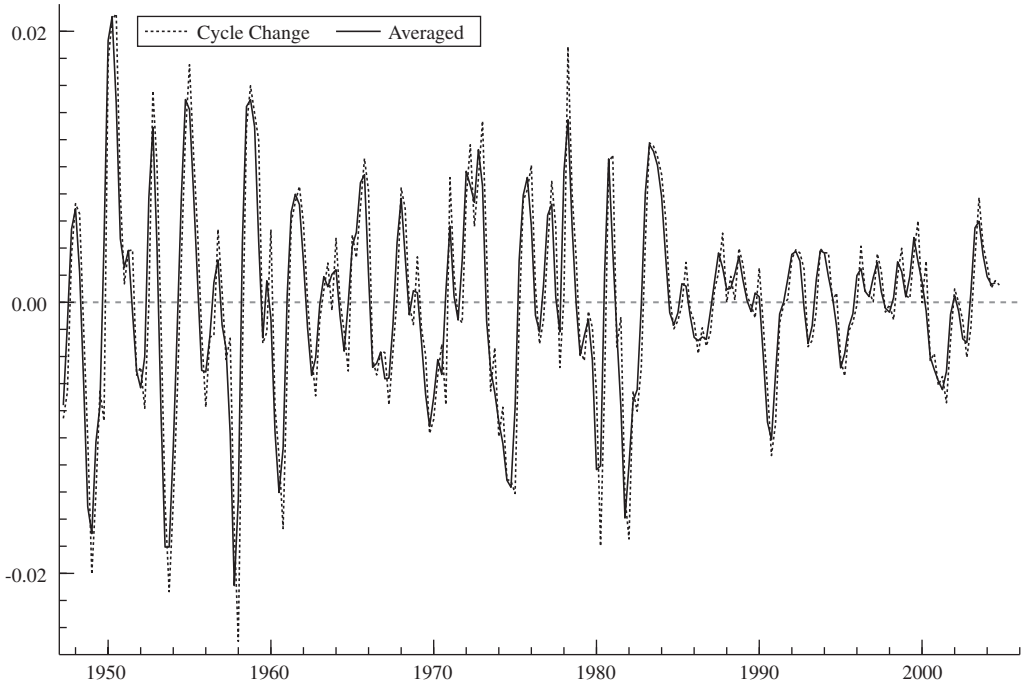

Fig. 10. $D\psi_{2,t}$ and average of adjacent differences.

Table 2
Probability that the rate of change in the cycle is negative for the period 1972Q1 to 1975Q4

Date	$\Pr(D\psi_{2,t}<0)$	Date	$\Pr(D\psi_{2,t}<0)$
1972Q1	0.0124	1974Q1	0.8536
1972Q2	0.1689	1974Q2	0.9924
1972Q3	0.0252	1974Q3	0.9954
1972Q4	0.0002	1974Q4	1.000
1973Q1	0.0438	1975Q1	0.8734
1973Q2	0.8114	1975Q2	0.1330
1973Q3	0.5200	1975Q3	0.0796
1973Q4	0.9426	1975Q4	0.0068

4.3. On line tracking of the output gap

While smoothed estimates for the whole series provide a historical perspective, what is most important for policy makers is real-time estimates of the current state of the economy. Fig. 11 tracks the filtered cycle from the first quarter of 1999 to the end of 2004. An associated series of the probability of the cycle being negative could also be produced.

Fig. 12 shows the estimated on-line change in the cycle based on filtered estimates of $D\psi_{2,t}$, while Fig. 13 shows the estimated probability that $D\psi_{2,t}$ is negative. As in Fig. 12,

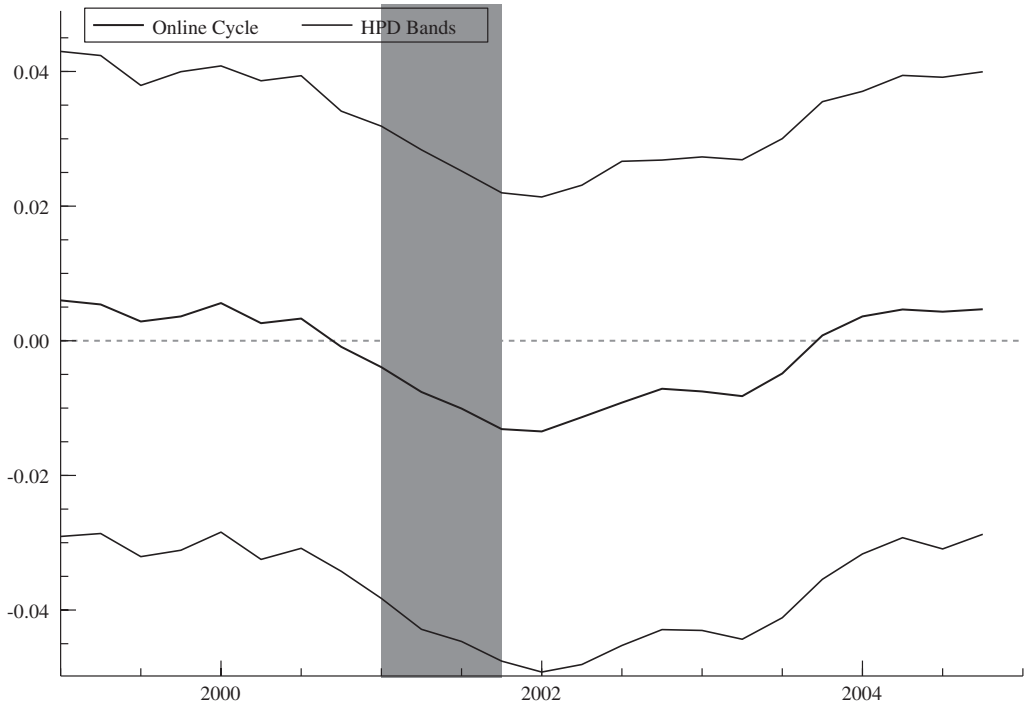

Fig. 11. Filtered cycle from 1999Q1 with 95% HPD bands.

these series clearly show the slowdown prior to the 2001 recession identified by the NBER; in particular, Fig. 13 indicates that the probability that the cycle is heading downwards in 2002 is between 0.65 and 0.70. An even more interesting feature of these graphs is that they show how $D\psi_{2,t}$ anticipates changes in the direction of the cycle. Thus, in 2002Q1 the cycle is still moving down but $D\psi_{2,t}$ is positive: in the next period the output gap has narrowed. Similar behavior can be seen in 2000Q3, 2002Q4 and 2003Q2.

Fig. 14 shows the filtered marginal posterior for the cycle at 2001Q4 together with some of the subsequent (smoothed) posteriors. The idea is to show how the uncertainty is reduced with the arrival of new data. The posterior is already much tighter after two periods. One year later, there is a sizeable gain in precision and by the end of 2003 the distribution is relatively compact and symmetric with a peak close to -0.01.

4.4. Multivariate models

In this section we investigate whether modeling GDP jointly with other US macroeconomic series can help to give a more accurate picture of the business cycle. The first sub-section fits a model of the form (1) to consumption, investment and GDP. Although most of the benefit to business cycle analysis is likely to come from the inclusion of investment, it is interesting to fit the model to the three series, firstly, to make a comparison with the well-known study of King et al. (1991) and secondly, to see whether

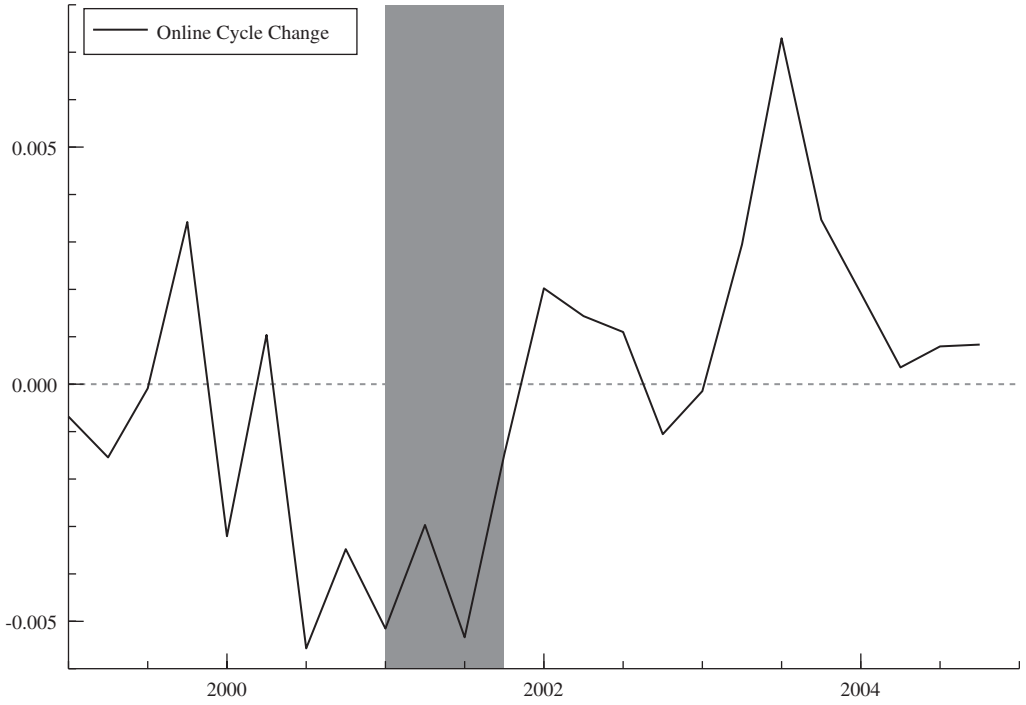

Fig. 12. Estimated change in the filtered cycle.

there are technical difficulties in applying our method in the trivariate case. The second sub-section reports an extension of the modeling framework in which lagged cyclical components are included in an inflation equation.

4.5. C–I–Y for the US economy

The data consist of quarterly series, expressed as real quantities with base year 2000, on personal consumption expenditures, gross private domestic investment, and GDP from 1947Q1 to 2004Q4 (source: Bureau of Economic Analysis). The similar cycle assumption of a common period is perfectly sensible, and it may lead to efficiency gains. Details of the extension of the MCMC routine to the general multivariate model, (1), are set out in the Appendix.

To demonstrate the effectiveness of pooling the data, we first assumed a flat prior on λ_c. (The uniform prior on $[0, 1]$ is again used for ρ.) To avoid potential distortions of the multivariate likelihood surface due to the influence of the inverted Wishart prior, we used flat priors on the variance matrices. The posterior density for λ_c with $n = 2$ reached a clear peak where the implied period is about 21 quarters.

If we use an informative prior for $n = 2$ there seems little point in going beyond the wide prior given the results of the previous paragraph—in any case the wide prior seemed to be adequate even in the univariate case. Table 3 shows posterior means of the variance parameters ($\times 10^7$), correlations between the components across different series (υ) and

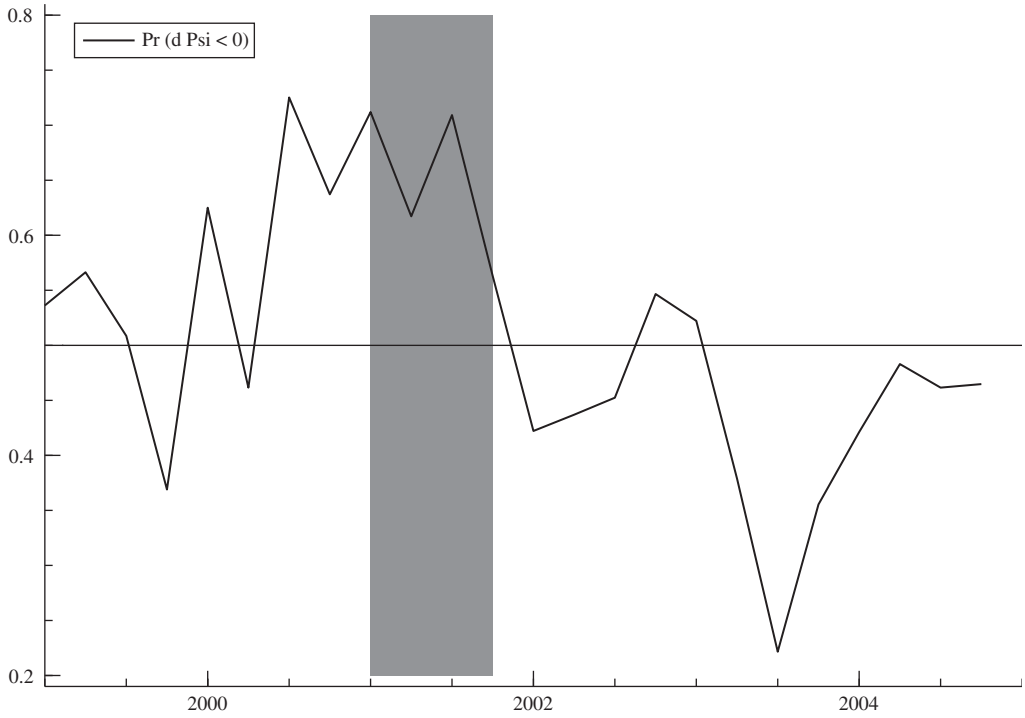

Fig. 13. Probability that change in the filtered cycle, $D\psi_{2,t}$, is negative.

shared cyclical parameters, ρ and λ_c. Random drawings for the correlations are directly constructed from the draws for the variance matrices.

The estimated second-order cycle in GDP is similar to the one produced by the univariate case. However, the HPD bands are approximately 25% smaller than the univariate band during the 1990s. The filtered estimates corresponding to those reported at the end of the previous section will likewise be more accurate.

Instead of looking at recent filtered estimates, as in the previous section, we examine forecasting performance. Fig. 15 shows the multistep predictions[7] for the cycle made at the end of 2002 for the subsequent two-year period. The forecast function indicates an upturn in the cycle and this is consistent with the smoothed estimates made at the end of 2004 (the corresponding univariate estimates shown in Fig. 8 are similar).

4.6. *Inflation and the output gap*

Kuttner (1994) and Planas and Rossi (2004) argue that the link between inflation and the output gap, as reflected in the Phillips curve, may be exploited to produce more reliable estimates of the output gap. Thus, an unobserved components model for GDP is combined with an equation in which inflation depends on lagged values of the output gap, as measured by the cycle in GDP.

[7]The subsequent observations are not being used to construct a series of one-step ahead predictions.

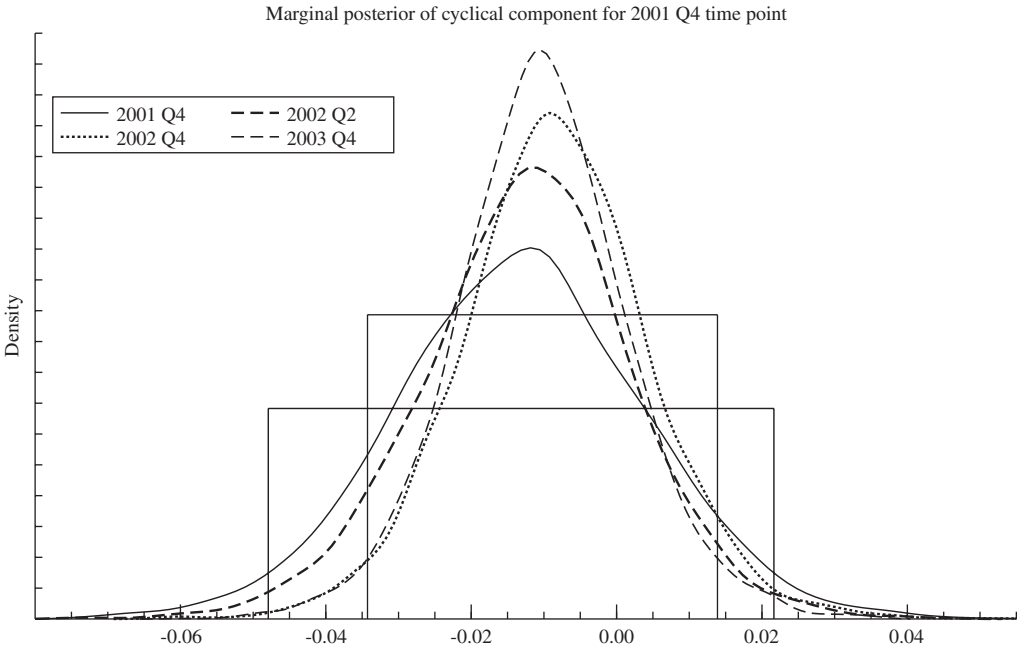

Fig. 14. Marginal posterior densities of $\psi_{2,t}$ (with a wide informative prior on λ_c) for 2001Q4 using data up to 2001Q4 and then using more observations. HPD regions are shown for the filtered estimate (at 2001Q4) and the smoothed estimate at the end of the series (2003Q4).

Table 3
Posterior means for a trivariate model for US GDP, consumption, and investment from 1947Q1 to 2004Q4 using flat priors on the variance matrices and a wide prior on λ_c

	Consumption			Investment			GDP		
n	σ_ζ^2	σ_κ^2	σ_ε^2	σ_ζ^2	σ_κ^2	σ_ε^2	σ_ζ^2	σ_κ^2	σ_ε^2
1	70.1	96.3	167	1640	7890	4430	106	279	82.1
2	40.5	99.6	193	338	7170	6160	51.8	260	137
3	39.4	64	208	296	4650	7221	66.6	168	169
4	40.9	33	218	300	1999	8220	86.5	75	191

Correlations and shared cyclical parameters

n	ρ	λ_c	$2\pi/\lambda_c$	υ_ζ	υ_ε	υ_κ	$m(y)$
1	0.886	0.545	11.6	0.927	−0.679	0.654	−1842.9
2	0.705	0.334	20.3	0.732	−0.493	0.824	−1724.9
3	0.603	0.266	34.6	0.709	−0.388	0.851	−1680.0
4	0.554	0.281	31.5	0.620	−0.345	0.895	−1663.1

The model we fit is

$$\begin{bmatrix} \pi_t \\ y_t \end{bmatrix} = \begin{bmatrix} \mu_t^\pi \\ \mu_t^y \end{bmatrix} + \begin{bmatrix} \psi_t^\pi \\ \psi_t^y \end{bmatrix} + \begin{bmatrix} p_t \\ 0 \end{bmatrix} + \begin{bmatrix} \varepsilon_t^\pi \\ \varepsilon_t^y \end{bmatrix},$$

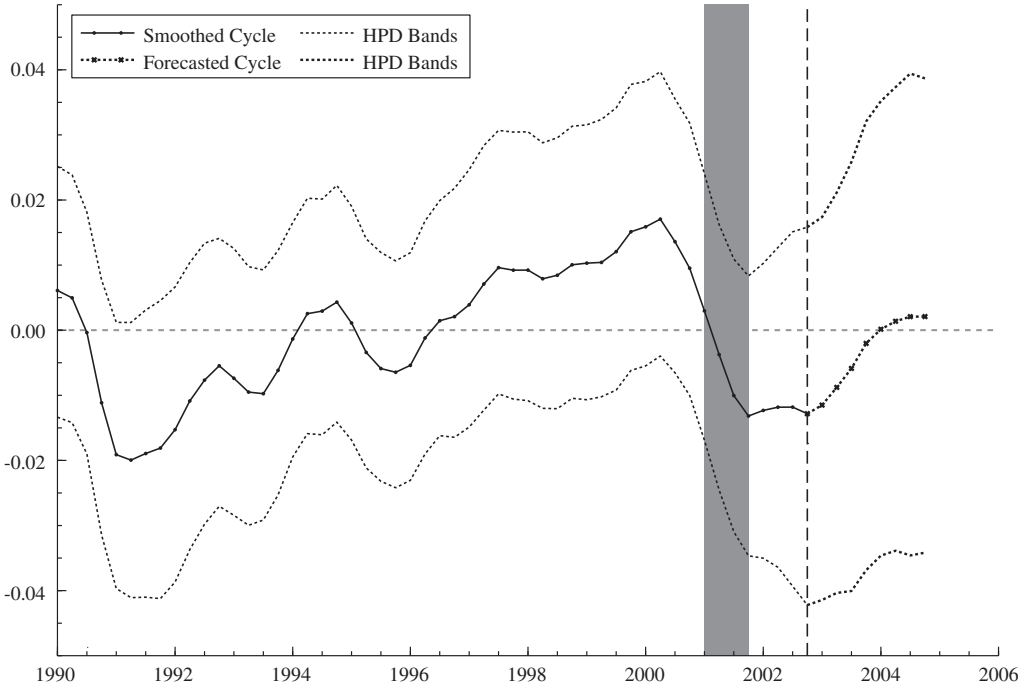

Fig. 15. Forecast of cyclical component, with 95% HPD bands, of US GDP for the two-year period 2003Q1–2004Q4, based on the trivariate C–I–Y model with $n = 2$ using a wide prior on λ_c. Also shown from 1990Q1 to 2002Q4 is the smoothed cycle.

where y_t is the logarithm of quarterly real US GDP, π_t is the CPI rate of inflation and

$$p_t = c_1\psi^y_{t-1} + c_2\psi^y_{t-2}$$

denotes the price pressure due to recent output gap levels. Note that a simple transformation of the vector $(\psi^\pi_t, \psi^y_t)'$ allows ψ^π_t to be broken down into two independent parts, one of which depends on the GDP cycle, that is $\psi^\pi_t = c\psi^y_t + \psi^+_t$. Putting the model into state space form is straightforward. Posterior results for cycle orders from 1 to 4 are shown in Table 4.

The above inflation equation is somewhat different from the ones in Kuttner (1994) and Planas and Rossi (2004) in that we drop the lagged growth rate of GDP and include a stochastic trend. The reason for including the stochastic trend is that it is difficult to find a stable relationship between inflation and output without it. Fig. 16 shows the trend and cycles obtained by a model of the form (1) with $n = 2$. Plotting the cycles on the same graph, Fig. 17, provides an indication that GDP leads inflation, particularly in the 1970s.

In moving from first- to second-order cycles, the marginal likelihood rises markedly, and the coefficients that relate the response of inflation to lagged output gap increase. The posterior distributions for c_1 and c_2 appear more or less symmetric with the 95% HPD bounds being [0.144, 1.380] and [−0.486, 0.470], respectively. Furthermore the correlation between the cycle disturbances is close to zero, implying a coefficient of only 0.027 on the

Table 4
Posterior means for a bivariate model of quarterly US CPI inflation and real GDP from 1947Q2 to 2004Q4 using flat priors on the variance matrices and a wide prior on the central frequency λ_c

	Inflation			GDP		
n	σ_ζ^2	σ_κ^2	σ_ε^2	σ_ζ^2	σ_κ^2	σ_ε^2
1	5.97	337	5241	44.4	497	39.6
2	3.77	112	5875	25.8	272	140
3	4.76	39.3	6315	62.6	131	173
4	4.25	19.3	6594	96.1	82	180

Correlations (υ), *shared cyclical parameters* (ρ *and* λ_c) *and inflation response coefficients* (c_1 *and* c_2)

n	ρ	λ_c	$2\pi/\lambda_c$	υ_ζ	υ_ε	υ_κ	c_1	c_2	$m(y)$
1	0.902	0.349	18.25	0.045	−0.008	0.016	0.710	0.061	−1136.0
2	0.725	0.295	22.36	−0.221	−0.028	0.040	0.723	0.018	−974.1
3	0.596	0.302	22.09	−0.081	−0.029	0.168	0.876	0.098	−948.2
4	0.501	0.313	21.54	0.105	−0.023	0.243	1.01	0.155	−944.5

All variance parameters are multiplied by 10^7.

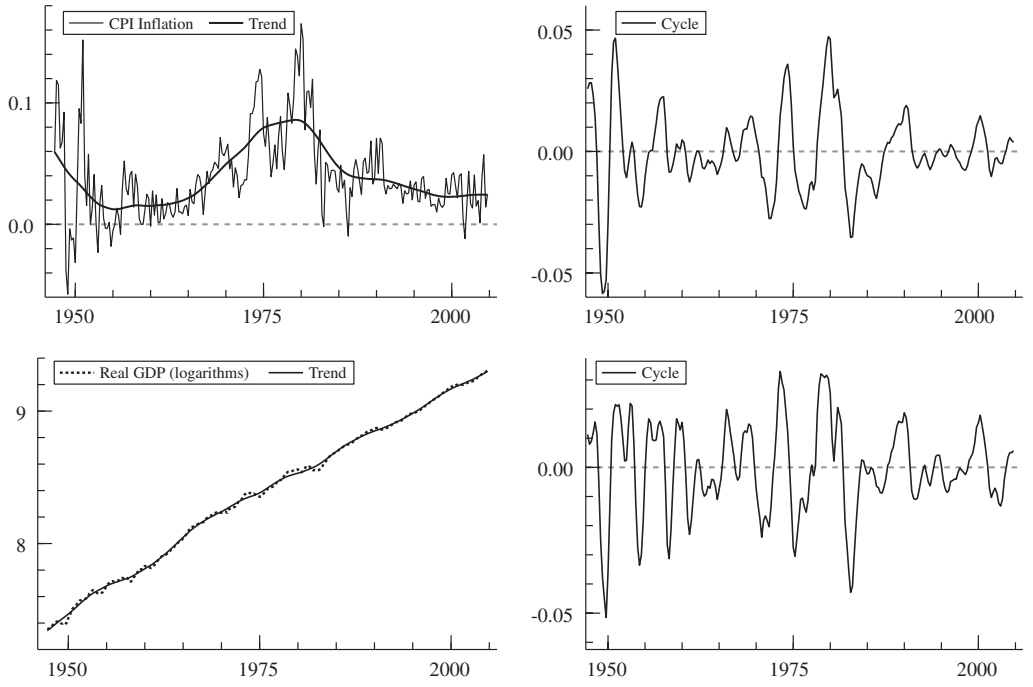

Fig. 16. Estimated trends and cycles in quarterly CPI inflation and US real GDP (logarithms) for $n = 2$ with wide prior on λ_c.

current cycle if the model is reparametrized in the way indicated earlier. The conclusion is that an increase in the output gap of one percent of GDP is expected to foreshadow a rise of about 0.7 of a percentage point in the rate of inflation in the next quarter. Including a

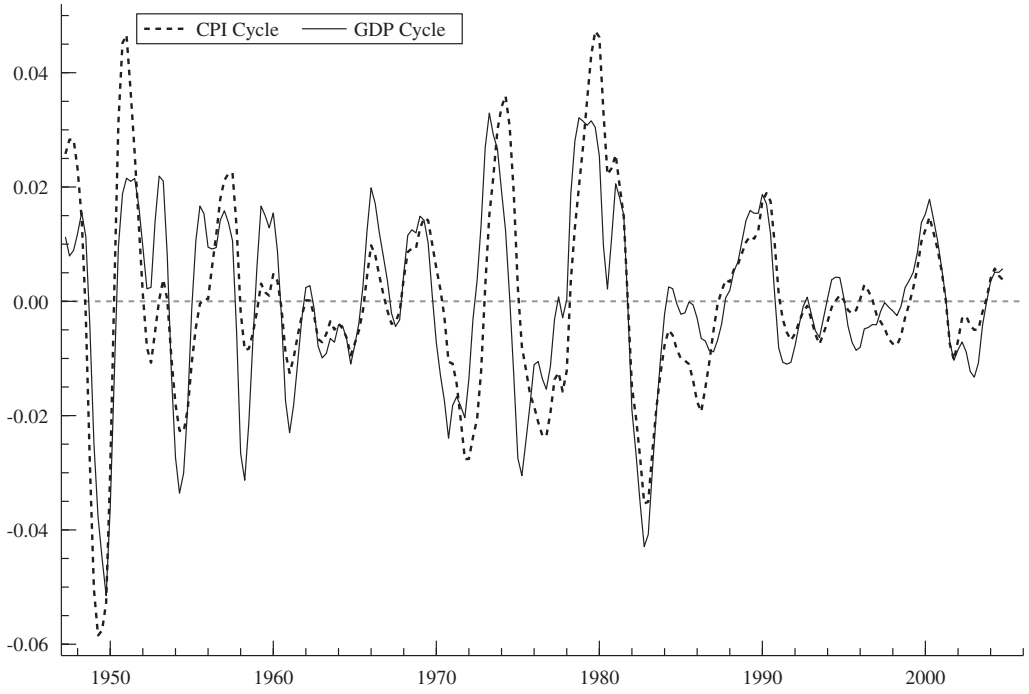

Fig. 17. Estimated second-order cycles in CPI inflation and US GDP.

second lag, as is done by Planas and Rossi (2004) but not by Kuttner (1994), is probably not necessary.

The GDP cycle resembles corresponding series from the univariate model, but the HPD bands are narrower. Of course, it would be even more useful if one had an equation that featured leads in the cycle rather than lags. Nevertheless given that it takes two or three periods to recognize a turning point, a one period lag is not entirely useless. Of course, the model might also be useful for forecasting inflation.

5. Conclusion

This article further extends the model-based methodology for the estimation of trends and cycles in macroeconomic time series. The preferred models use the second-order cycles introduced recently by Harvey and Trimbur (2003) since these tend to be smoother than first-order cycles, with relatively more noise consigned to the irregular component. We suggest various ways in which the information obtained by fitting such models can be used to describe past movements of the cycle and to focus attention on such features such as changing volatility and turning points. Handling several cycles, such as minor inventory (Abramovitz) cycles, construction cycles or Kondratieff long swings, is also possible within our framework.[8]

[8]There is an illustrative example (concerned with rainfall) in the STAMP manual of Koopman et al. (2000, pp. 55–56, see also exercise 5, p. 57), see also Harvey et al. (2005).

There are two main attractions to a Bayesian approach. Firstly, flexible restrictions can be placed on key parameters, such as the frequency parameter in the stochastic cycle, and this avoids fitting implausible models. Secondly, parameter uncertainty is taken account of in providing information about extracted components and their forecasts. The disadvantage is that the computational requirements are heavier than for ML. However, although Bayesian estimation typically requires several minutes,[9] as opposed to a few seconds, this hardly renders it infeasible. The MCMC routines we describe should be of value for future research, particularly for multivariate series where the computations are nontrivial.

Cycles were successfully extracted from US GDP using a univariate model. Such cycles have a simple interpretation in terms of the percentage by which they exceed or fall below the long-term level of potential output. Associated measures track the size and direction of the cycle. The posterior distributions clearly indicate the degree of uncertainty that arises from signal extraction and the fact that the parameters are unknown. Fitting multivariate models offers the possibility of reducing this uncertainty. Used on-line, measures such as the probability that the output gap is increasing or decreasing may be of considerable practical value.

Acknowledgments

Earlier versions of this paper were presented at the Cambridge PhD workshop, the NAKE Conference, ESAM2004 and JRC in Italy. Trimbur wishes to thank the Cambridge Commonwealth Trust and the Richard Kahn Fund for financial support, and the Tinbergen Institute for its hospitality and financial support in autumn 2001. He is also grateful for support as a Post-Doctoral Research Fellow at the US Census Bureau. 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. Van Dijk thanks the Erasmus Research Institute for Management for financial assistance. We are grateful to Bill Bell, Simon Godsill, Neil Shephard, Arnold Zellner, an Associate Editor and two referees for helpful comments.

Appendix A. Rate of change in the cycle

A first-order cycle in continuous time is written as

$$\mathrm{d}\boldsymbol{\psi}(t) = \mathbf{A}\boldsymbol{\psi}(t)\,\mathrm{d}t + \mathrm{d}\mathbf{W}_{\kappa}(t),$$

where $\boldsymbol{\psi}(t) = (\psi(t)\ \psi^{*}(t))', \mathbf{W}_{\kappa}(t)$ is a 2×1 vector of Brownian motion and

$$\mathbf{A} = \begin{bmatrix} \log\rho & \lambda_c \\ -\lambda_c & \log\rho \end{bmatrix}.$$

[9] Univariate: 5000 iterations takes about 40 s for $n = 2$ and about 80 s for $n = 4$. Output gap/bivariate: 5000 iterations takes 4 min for $n = 2$. Trivariate C–I–Y: 5000 iterations takes about 5 min, 45 s for $n = 2$. Recall that when we take 5000 draws with 10 iterations per draw and 10,000 burned there are 600,000 iterations. Microprocessor of computer is Pentium 4 (2.26 GHz).

The expected incremental change in $\psi(t)$ is

$$\mathrm{d}\psi(t) = (\log \rho)\psi(t)\,\mathrm{d}t + \lambda_c \psi^*(t)\,\mathrm{d}t$$

and the discrete time model is as in (3); see Harvey (1989, p. 487). The discrete time expression corresponding to $\mathrm{d}\psi(t)$ is therefore as in (10). There is a slight complication with flow variables in that the discrete time model has correlated measurement and transition equation noise; see Harvey (1989, p. 494). However, we would still suggest using (10) as an approximation.

The second-order cycle is

$$\mathrm{d}\boldsymbol{\psi}_2(t) = \mathbf{A}\boldsymbol{\psi}_2(t)\,\mathrm{d}t + \boldsymbol{\psi}_1(t)\,\mathrm{d}t,$$
$$\mathrm{d}\boldsymbol{\psi}_1(t) = \mathbf{A}\boldsymbol{\psi}_1(t)\,\mathrm{d}t + \mathrm{d}\mathbf{W}_\kappa(t)$$

with expected incremental change

$$\mathrm{d}\psi_2(t) = (\log \rho)\psi_2(t)\,\mathrm{d}t + \lambda_c \psi_2^*(t)\,\mathrm{d}t + \psi_1(t)\,\mathrm{d}t. \tag{14}$$

The model in (4), that is

$$\boldsymbol{\psi}_{2,t} = \mathbf{T}\boldsymbol{\psi}_{2,t-1} + \boldsymbol{\psi}_{1,t-1}, \tag{15}$$

$$\boldsymbol{\psi}_{1,t} = \mathbf{T}\boldsymbol{\psi}_{1,t-1} + \boldsymbol{\kappa}_t, \quad \boldsymbol{\kappa}_t \sim NID(\mathbf{0}, \sigma_\kappa^2 \mathbf{I}) \tag{16}$$

with $\mathbf{T}$ as in (7) and $\boldsymbol{\psi}_{1,t} = \mathbf{T}\boldsymbol{\psi}_{1,t}^{\dagger}$, where $\boldsymbol{\psi}_{1,t}^{\dagger}$ corresponds to the continuous time variable $\boldsymbol{\psi}_1(t)$, is an approximation to the discrete time formulation as there will be a disturbance, correlated with $\boldsymbol{\kappa}_t$, attached to the first equation and, for a flow, both disturbances will be correlated with the measurement equation disturbance. If the approximation to the continuous time model is accepted, the appropriate formula for change is as in (11) because of the transformation from $\boldsymbol{\psi}_{1,t}^{\dagger}$ to $\boldsymbol{\psi}_{1,t}$.

Appendix B. Priors

The informative priors on λ_c are constructed from the class of beta distributions. This gives a compact and flexible way to specify expectations of business cycle periods and their associated uncertainty. For details on the beta probability distribution, we refer to Poirier (1995).

A mode of $2\pi/20$ is used for quarterly US macroeconomic data. This can be easily adjusted to implement alternative expectations. In the current application, the mode of $2\pi/20$ reflects a central business cycle period of five years. Further, the prior is constrained to lie on the interval $[\pi/20, \pi/4]$; this preempts average periods greater than ten years and less than two years. Concentrating on the mode ensures that the prior probability density is maximum in the desired region.

The prior design makes it easy to adjust the implied uncertainty in the expectations. The dispersion of the prior is controlled by the beta-distribution parameters, that may be set to yield the appropriate standard deviation. The three priors in Fig. 1 range from weakly to highly informative. For the widest prior the standard deviation is $\sigma_\lambda = 2\pi/50$ and the proportional spread is $\sigma_\lambda/\widehat{\mu}_\lambda = 40\%$. For the intermediate prior $\sigma_\lambda = 2\pi/150, \sigma_\lambda/\widehat{\mu}_\lambda = 13\%$, and for the sharpest prior $2\pi/400, \sigma_\lambda/\widehat{\mu}_\lambda = 5\%$.

The flat priors on the variance parameters $\sigma_\zeta^2, \sigma_\kappa^2, \sigma_\varepsilon^2$ ensure that the posterior is unaffected by the prior shape around zero. As shown below, the variance parameters have independent inverted gamma conditional posteriors when independent flat priors are used,

so direct simulation is possible, which enhances the efficiency of the algorithm. For details on the inverted gamma family we refer to Poirier (1995).

Another possibility would be to use inverted gamma densities. These are natural conjugate priors so the conditional posteriors would again be inverted gamma. This would allow for informative priors on the variance parameters. However, the use of inverted gamma densities with very low shape and scale, to represent vague prior notions, can lead to problems in estimation and an inaccurate assessment. Specifically, the posterior may be significantly influenced by the steep descent of the prior around zero; the net effect depends on the behavior of the likelihood function near the origin. For parameters that tend to take on small values, such as the variance of the slope disturbance in the stochastic trend, the use of the flat prior is recommended to avoid distortions. The exact choice of the interval is, in our analysis, less important as it can be set relatively large to include any plausible prior value. For a similar argument in favor of uniform priors we refer to Gelman (2006).

Similarly, in the multivariate case, we work with flat priors defined over an arbitrarily large N-dimensional region. An alternative would be to use natural conjugate inverted Wishart priors. These are the generalization of the inverted gamma to the multivariate setting; see Zellner (1971), for instance. However, the use of a limiting inverted Wishart density to represent a noninformative prior on a variance matrix involves the same difficulties as in the univariate setting, and the possible distortions in the likelihood surface may be more complex (the number of implicit variance and covariance parameters increases rapidly with the number of series being modeled). Of course, an informative inverted Wishart could be used, but even when there are expectations about correlations across series, it is unclear how to express such prior knowledge in the formulation of the inverted Wishart density. In any case, in using flat priors on the variance matrices $\Sigma_\kappa, \Sigma_\zeta, \Sigma_\varepsilon$, the inverted Wishart form for the conditional posteriors continues to hold.

Appendix C. MCMC routine

This appendix describes the algorithm used for posterior analysis of the class of models (1). The MCMC routine, constructed a 'Metropolis–Hastings within Gibbs' sampling algorithm, has a design that can be naturally extended to handle other model structures. The algorithm generates pseudo-random drawings from the combined posterior density of the components and parameter vector; by taking suitable transformations, properties of marginal and joint densities are studied and summary measures estimated. We set out the algorithm in the general multivariate case and then comment on the special case of a univariate model and on the use of variance priors.

The joint posterior $p(\boldsymbol{\theta}, \boldsymbol{\alpha}|\mathbf{y})$ is the target density. We consider the state matrix $\boldsymbol{\alpha}$ as a single block and sub-divide the parameter vector $\boldsymbol{\theta}$ into the set of variance parameters, along with ρ and λ_c taken individually. This gives a total of four blocks in the Gibbs sampler. The corresponding steps at each iteration of the sampling algorithm are described below.

After choosing a set of initial values $\boldsymbol{\theta}^{(0)}$, the MCMC routine produces a sequence of draws $\{\boldsymbol{\theta}^{(i)}, \boldsymbol{\alpha}^{(i)}\}$, $i = 1, 2, \ldots$, according to the following four steps:

1. $\boldsymbol{\alpha}^{(i)}$ is drawn from $p(\boldsymbol{\alpha}|\boldsymbol{\theta}^{(i-1)}, \mathbf{y})$.
2. $\boldsymbol{\Sigma}_\kappa^{(i)}, \boldsymbol{\Sigma}_\zeta^{(i)}, \boldsymbol{\Sigma}_\varepsilon^{(i)}$ are collectively sampled from $p(\boldsymbol{\Sigma}_\kappa, \boldsymbol{\Sigma}_\zeta, \boldsymbol{\Sigma}_\varepsilon|\rho^{(i-1)}, \lambda_c^{(i-1)}, \boldsymbol{\alpha}^{(i)}, \mathbf{y})$.

3. $\rho^{(i)}$ is drawn from $p(\rho|\boldsymbol{\Sigma}_\kappa^{(i)}, \boldsymbol{\Sigma}_\zeta^{(i)}, \boldsymbol{\Sigma}_\varepsilon^{(i)}, \lambda_c^{(i-1)}, \boldsymbol{\alpha}^{(i)}, \mathbf{y})$.
4. $\lambda_c^{(i)}$ is drawn from $p(\lambda_c|\boldsymbol{\Sigma}_\kappa^{(i)}, \boldsymbol{\Sigma}_\zeta^{(i)}, \boldsymbol{\Sigma}_\varepsilon^{(i)}, \rho^{(i)}, \boldsymbol{\alpha}^{(i)}, \mathbf{y})$.

We now describe how the draw from the conditional density is obtained at each step. For details on the matrix calculus used in certain parts of the proof, we refer to Magnus and Neudecker (1999).

1. $p(\boldsymbol{\alpha}|\boldsymbol{\theta}^{(i-1)}, \mathbf{y})$: Eqs. (1)–(4) express the observed process as a sum of unobserved components. The state space form of the model is set out in Trimbur (2006). There are $(2+2n)N$ elements in the state vector, including the trend, the cycle, and the processes used in their definition:

$$\boldsymbol{\alpha}_t = [\boldsymbol{\mu}_t', \boldsymbol{\beta}_t', \boldsymbol{\psi}_t']', \tag{17}$$

$$\boldsymbol{\psi}_t = [\boldsymbol{\psi}_{n,t} \;\; \boldsymbol{\psi}_{n,t}^* \;\; \ldots \;\; \boldsymbol{\psi}_{1,t} \;\; \boldsymbol{\psi}_{1,t}^*]'. \tag{18}$$

This gives a partition of the state vector into the components linked to the stochastic trend, $\boldsymbol{\mu}_t$, and the components linked to the stochastic cycle, $\boldsymbol{\psi}_{n,t}$. The vector of length $(2+2n)NT$ formed by stacking the complete set of state vectors over the sample period is denoted as $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1', \ldots, \boldsymbol{\alpha}_T']'$. The measurement equation in the state space form can be written as

$$\mathbf{y}_t = \mathbf{z}_t\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, T,$$

where $\mathbf{z}_t$ is the appropriate selection matrix of ones and zeros.

Given the most recent iterate of the parameter vector $\boldsymbol{\theta}^{(i-1)}$, the general simulation smoother of Durbin and Koopman (2002) is applied to generate a draw $\boldsymbol{\alpha}^{(i)}$ from $p(\boldsymbol{\alpha}|\boldsymbol{\theta}^{(i-1)}, \mathbf{y})$.

2. $p(\boldsymbol{\Sigma}_\kappa, \boldsymbol{\Sigma}_\zeta, \boldsymbol{\Sigma}_\varepsilon|\rho^{(i-1)}, \lambda_c^{(i-1)}, \boldsymbol{\alpha}^{(i)}, \mathbf{y})$. With independent priors,

$$p(\boldsymbol{\Sigma}_\kappa, \boldsymbol{\Sigma}_\zeta, \boldsymbol{\Sigma}_\varepsilon|\rho, \lambda_c, \boldsymbol{\alpha}, \mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\theta})p(\boldsymbol{\alpha}|\boldsymbol{\theta})p(\boldsymbol{\Sigma}_\kappa)p(\boldsymbol{\Sigma}_\zeta)p(\boldsymbol{\Sigma}_\varepsilon). \tag{19}$$

The joint density of the set of state vectors, given the parameters, is

$$p(\boldsymbol{\alpha}|\boldsymbol{\theta}) = p(\boldsymbol{\alpha}_1|\boldsymbol{\theta}) \prod_{t=2}^{T} p(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}, \boldsymbol{\theta}). \tag{20}$$

The above expression involves the initial distribution $p(\boldsymbol{\alpha}_1|\boldsymbol{\theta}) = p(\boldsymbol{\mu}_1', \boldsymbol{\beta}_1', \boldsymbol{\psi}_1'|\boldsymbol{\theta}) = p(\boldsymbol{\mu}_1', \boldsymbol{\beta}_1'|\boldsymbol{\theta}) \times p(\boldsymbol{\psi}_1'|\boldsymbol{\theta})$. Following standard practice, it is assumed that $p(\boldsymbol{\mu}_1', \boldsymbol{\beta}_1'|\boldsymbol{\theta})$ is diffuse. Specifically, it is a mean-zero Gaussian density with $var([\boldsymbol{\mu}_1', \boldsymbol{\beta}_1']') = \sigma_*^2 I_{2N}$, where $\sigma_*^2 \to \infty$. The exact initial conditions for the cyclical state $\boldsymbol{\psi}_1'$ are given in Trimbur (2006).

While some elements of $\boldsymbol{\alpha}_t$ are deterministic functions of $\boldsymbol{\alpha}_{t-1}$, the conditional density of $\boldsymbol{\alpha}_t$ given $\boldsymbol{\alpha}_{t-1}$ depends only on the stochastic elements. To focus on the relevant subset of the state space, we define the reduced state vector, transition matrix, and covariance matrix as follows:

$$\boldsymbol{\alpha}_t^* = \begin{bmatrix} \boldsymbol{\beta}_t \\ \boldsymbol{\psi}_{1,t} \\ \boldsymbol{\psi}_{1,t}^* \end{bmatrix}, \quad \mathbf{T}^* = \begin{bmatrix} \mathbf{I}_N & \mathbf{O} \\ \mathbf{O} & \mathbf{T} \otimes \mathbf{I}_N \end{bmatrix}, \quad \boldsymbol{\Omega}^* = \begin{bmatrix} \boldsymbol{\Sigma}_\zeta & \mathbf{O} \\ 0 & \mathbf{I}_2 \otimes \boldsymbol{\Sigma}_\kappa \end{bmatrix}.$$

The one-step ahead density is

$$p(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}, \boldsymbol{\theta}) = |\boldsymbol{\Omega}^*|^{-1/2} \exp\{-\tfrac{1}{2}(\boldsymbol{\alpha}_t^* - \mathbf{T}^*\boldsymbol{\alpha}_{t-1}^*)'\boldsymbol{\Omega}^{*-1}(\boldsymbol{\alpha}_t^* - \mathbf{T}^*\boldsymbol{\alpha}_{t-1}^*)\}.$$

Denoting the variance matrix of the cyclical state vector as $\boldsymbol{\Gamma} = var(\boldsymbol{\psi}_t)$, we have

$$p(\boldsymbol{\alpha}|\boldsymbol{\theta}) \propto |\boldsymbol{\Gamma}|^{-1/2}|\boldsymbol{\Omega}^*|^{-(T-1)/2} \times \exp\left\{-\frac{1}{2}\boldsymbol{\psi}_1'\boldsymbol{\Gamma}^{-1}\boldsymbol{\psi}_1 - \frac{1}{2}\sum_{t=2}^{T}(\boldsymbol{\alpha}_t^* - \mathbf{T}^*\boldsymbol{\alpha}_{t-1}^*)'\boldsymbol{\Omega}^{*-1}(\boldsymbol{\alpha}_t^* - \mathbf{T}^*\boldsymbol{\alpha}_{t-1}^*)\right\}.$$

This can be rewritten as

$$\begin{aligned} p(\boldsymbol{\alpha}|\boldsymbol{\Sigma}_\zeta, \boldsymbol{\Sigma}_\kappa, \rho, \lambda_c) \propto |\boldsymbol{\Gamma}|^{-1/2}|\mathbf{I}_2 \otimes \boldsymbol{\Sigma}_\kappa|^{-(T-1)/2}|\boldsymbol{\Sigma}_\zeta|^{-(T-1)/2} \\ \times \exp\left\{-\frac{1}{2}\boldsymbol{\psi}_1\boldsymbol{\Gamma}^{-1}\boldsymbol{\psi}_1 - \frac{1}{2}\sum_{t=2}^{T}\mathbf{c}_t'(\mathbf{I}_2 \otimes \boldsymbol{\Sigma}_\kappa)^{-1}\mathbf{c}_t\right\} \\ \times \exp\left\{-\frac{1}{2}\sum_{t=2}^{T}(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t-1})'\boldsymbol{\Sigma}_\zeta^{-1}(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t-1})\right\}, \end{aligned} \tag{21}$$

where $\mathbf{c}_t = [\boldsymbol{\psi}_{1,t}', \boldsymbol{\psi}_{1,t}^{*'}]' - (\mathbf{T} \otimes \mathbf{I}_N)[\boldsymbol{\psi}_{1,t-1}', \boldsymbol{\psi}_{1,t-1}^{*'}]'$ with $\boldsymbol{\psi}_{1,t} = [\psi_{1,t}^1, \ldots, \psi_{1,t}^N]'$, $\boldsymbol{\psi}_{1,t}^* = [\psi_{1,t}^{1*}, \ldots, \psi_{1,t}^{N*}]'$.

The summation in the cyclical part is

$$\begin{aligned} \sum_{t=2}^{T}\mathbf{c}_t'(\mathbf{I}_2 \otimes \boldsymbol{\Sigma}_\kappa)^{-1}\mathbf{c}_t &= \sum_{t=2}^{T}\mathrm{tr}[(\mathbf{I}_2 \otimes \boldsymbol{\Sigma}_\kappa^{-1})\mathbf{c}_t\mathbf{c}_t'] \\ &= \mathrm{tr}(\boldsymbol{\Sigma}_\kappa^{-1}\mathbf{c}_{1,t}\mathbf{c}_{1,t}') + \mathrm{tr}(\boldsymbol{\Sigma}_\kappa^{-1}\mathbf{c}_{2,t}\mathbf{c}_{2,t}'), \end{aligned}$$

where $\mathbf{c}_t' = [\mathbf{c}_{1,t}' \ \mathbf{c}_{2,t}']$. For this partition of $\mathbf{c}_t$, the upper half $\mathbf{c}_{1,t}$ corresponds to the N first-order cycles $[\psi_{1,t}^1, \ldots, \psi_{1,t}^N]'$, and the lower half $\mathbf{c}_{2,t}$ to the auxiliaries, $[\psi_{1,t}^{*1}, \ldots, \psi_{1,t}^{*N}]'$. Further, it is shown in Trimbur (2006) that the unconditional variance matrix has the form $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}^* \otimes \boldsymbol{\Sigma}_\kappa$. Therefore, in the conditional density of $\boldsymbol{\alpha}$, the cyclical part of the model gives the factor

$$|\boldsymbol{\Gamma}^* \otimes \boldsymbol{\Sigma}_\kappa|^{-1/2}|\mathbf{I}_2 \otimes \boldsymbol{\Sigma}_\kappa|^{-(T-1)/2} \times \exp\left\{-\tfrac{1}{2}\boldsymbol{\psi}_1'[\boldsymbol{\Gamma}^* \otimes \boldsymbol{\Sigma}_\kappa]^{-1}\boldsymbol{\psi}_1 - \tfrac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}_\kappa^{-1}\mathbf{G})\right\},$$

where

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

Since

$$|\boldsymbol{\Gamma}^* \otimes \boldsymbol{\Sigma}_\kappa| = |\boldsymbol{\Gamma}^*|^N|\boldsymbol{\Sigma}_\kappa|^{2n}$$

and $|\mathbf{I}_2 \otimes \boldsymbol{\Sigma}_\kappa| = |\boldsymbol{\Sigma}_\kappa|^2$, it follows that

$$|\boldsymbol{\Gamma}^* \otimes \boldsymbol{\Sigma}_\kappa|^{-1/2}|\mathbf{I}_2 \otimes \boldsymbol{\Sigma}_\kappa|^{-(T-1)/2} = |\boldsymbol{\Gamma}^*|^{-N/2}|\boldsymbol{\Sigma}_\kappa|^{-(n+T-1)}.$$

Let the elements of block i,j of $\boldsymbol{\Gamma}^*$ be denoted by

$$\{\boldsymbol{\Gamma}^*\}_{i,j} = \begin{bmatrix} \boldsymbol{\Gamma}_{i,j}^* & \boldsymbol{\Gamma}_{i,j*}^* \\ \boldsymbol{\Gamma}_{i*,j}^* & \boldsymbol{\Gamma}_{i*,j*}^* \end{bmatrix}.$$

The complete cyclical state vector at each t is partitioned as

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

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

$$\boldsymbol{\psi}'_1[\boldsymbol{\Gamma}^*\otimes\boldsymbol{\Sigma}_\kappa]^{-1}\boldsymbol{\psi}_1 = \tfrac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}_\kappa^{-1}\mathbf{H}), \tag{22}$$

where the $(N\times N)$ matrix $\mathbf{H}$ is given by

$$\begin{aligned}\mathbf{H} = &\sum_{i=1}^{n}\left(\sum_{j=1}^{n}\boldsymbol{\Gamma}^{*-1}_{i,j}\boldsymbol{\psi}_{j,1}\boldsymbol{\psi}'_{i,1}+\sum_{j=1}^{n}\boldsymbol{\Gamma}^{*-1}_{i,j*}\boldsymbol{\psi}^*_{j,1}\boldsymbol{\psi}'_{i,1}\right)\\ &+\sum_{i=1}^{n}\left(\sum_{j=1}^{n}\boldsymbol{\Gamma}^{*-1}_{i*,j}\boldsymbol{\psi}_{j,1}\boldsymbol{\psi}^{*'}_{i,1}+\sum_{j=1}^{n}\boldsymbol{\Gamma}^{*-1}_{i*,j*}\boldsymbol{\psi}^*_{j,1}\boldsymbol{\psi}^{*'}_{i,1}\right).\end{aligned}$$

This expression is derived by first noting that

$$[\boldsymbol{\Gamma}^*\otimes\boldsymbol{\Sigma}_\kappa]^{-1} = \mathrm{tr}[(\boldsymbol{\Gamma}^{*-1}\otimes\boldsymbol{\Sigma}_\kappa^{-1})\boldsymbol{\psi}_1\boldsymbol{\psi}'_1].$$

Once the Kronecker product and multiplication are applied, the sum of the diagonal elements gives (22).

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

$$\sum_{t=2}^{T}(\boldsymbol{\beta}_t-\boldsymbol{\beta}_{t-1})'\boldsymbol{\Sigma}_\zeta^{-1}(\boldsymbol{\beta}_t-\boldsymbol{\beta}_{t-1}) = \sum_{t=2}^{T}\mathrm{tr}[\boldsymbol{\Sigma}_\zeta^{-1}(\boldsymbol{\beta}_t-\boldsymbol{\beta}_{t-1})(\boldsymbol{\beta}_t-\boldsymbol{\beta}_{t-1})'].$$

The conditional density of $\boldsymbol{\alpha}$ is therefore

$$\begin{aligned}p(\boldsymbol{\alpha}|\boldsymbol{\Sigma}_\zeta,\boldsymbol{\Sigma}_\kappa,\rho,\lambda_c) \propto\ &|\boldsymbol{\Gamma}^*|^{-N/2}|\boldsymbol{\Sigma}_\kappa|^{-(n+T-1)}\exp\left[-\tfrac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}_\kappa^{-1}[\mathbf{G}+\mathbf{H}])\right]\\ &\times|\boldsymbol{\Sigma}_\zeta|^{-(T-1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}_\zeta^{-1}\left[\sum_{t=2}^{T}(\boldsymbol{\beta}_t-\boldsymbol{\beta}_{t-1})(\boldsymbol{\beta}_t-\boldsymbol{\beta}_{t-1})'\right]\right)\right\}.\end{aligned} \tag{23}$$

The joint density of the observations given the hyperparameters and $\boldsymbol{\alpha}$ is

$$p(\mathbf{y}|\boldsymbol{\alpha},\boldsymbol{\theta}) \propto |\boldsymbol{\Sigma}_\varepsilon|^{-T/2}\exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\mathrm{tr}[\boldsymbol{\Sigma}_\varepsilon^{-1}(\boldsymbol{y}_t-\mathbf{z}_t\boldsymbol{\alpha}_t)(\boldsymbol{y}_t-\mathbf{z}_t\boldsymbol{\alpha}_t)']\right\}.$$

The use of a flat prior for $\boldsymbol{\Sigma}_\kappa$ means that $p(\boldsymbol{\Sigma}_\kappa)$ is proportional to a small, positive constant on its domain. The domain for $\boldsymbol{\Sigma}_\kappa$ covers the region of permissible values, which includes symmetric, positive semi-definite matrices of dimension N. Naturally, a flat prior covering the unbounded space of all permissible covariance matrices can only be defined as a limiting case. However, in practice it is sufficient to define the prior on an arbitrary finite sub-space; one can then choose a sub-space large enough to include all viable $\boldsymbol{\Sigma}_\kappa$. A similar assumption is made in the design of $p(\boldsymbol{\Sigma}_\zeta)$ and $p(\boldsymbol{\Sigma}_\varepsilon)$. In this framework, the conditional posterior in (19) is then proportional to the product of the conditional densities of $\boldsymbol{\alpha}$ and $\mathbf{y}$ shown above.

It follows that (19) factors into three distinct kernels of the inverted Wishart form. Therefore, in Step 2 of the Gibbs sampler, the conditional posterior for the variance

matrices $\boldsymbol{\Sigma}_\zeta$, $\boldsymbol{\Sigma}_\kappa$, $\boldsymbol{\Sigma}_\varepsilon$ consists of three independent inverted Wishart distributions. The shape parameter is constant for each density, as indicated below, while the scale matrix is updated at each iteration:

$$\boldsymbol{\Sigma}_\zeta : c_\zeta^* = c_\zeta + (T-1), \quad \mathbf{S}_\zeta = \sum_{t=2}^{T} (\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t-1})(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t-1})',$$

$$\boldsymbol{\Sigma}_\kappa : c_\kappa^* = c_\kappa + 2(n+T-1), \quad \mathbf{S}_\kappa = \mathbf{G}^{(i)} + \mathbf{H}^{(i)},$$

$$\boldsymbol{\Sigma}_\varepsilon : c_\varepsilon^* = c_\varepsilon + T, \quad \mathbf{S}_\varepsilon = \sum_{t=1}^{T} (\mathbf{y}_t - \mathbf{z}_t\boldsymbol{\alpha}_t^{(i)})(\mathbf{y}_t - \mathbf{z}_t\boldsymbol{\alpha}_t^{(i)})'.$$

The superscript in $\mathbf{G}^{(i)}$ means that the quantity is computed from the state vector drawn in the previous step of the same iteration, together with the cycle parameters from the previous iteration. At iteration i, the collective draw $\{\boldsymbol{\Sigma}_\zeta^{(i)}, \boldsymbol{\Sigma}_\kappa^{(i)}, \boldsymbol{\Sigma}_\varepsilon^{(i)}\}$ is generated by separately drawing from each inverted Wishart density.

3. $p(\rho|\boldsymbol{\Sigma}_\zeta^{(i)}, \boldsymbol{\Sigma}_\kappa^{(i)}, \boldsymbol{\Sigma}_\varepsilon^{(i)}, \lambda_c^{(i-1)}, \boldsymbol{\alpha}^{(i)}, \mathbf{y})$: The parameter ρ controls the persistence of cyclical shocks and how they are dampened over time. We use a uniform prior on $[0, 1]$ to ensure that the model produces a well-defined cycle. Using Bayes theorem for densities,

$$p(\rho|\boldsymbol{\Sigma}_\zeta, \boldsymbol{\Sigma}_\kappa, \boldsymbol{\Sigma}_\varepsilon, \lambda_c, \boldsymbol{\alpha}, \mathbf{y}) \propto p(\rho)p(\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\theta})p(\boldsymbol{\alpha}|\boldsymbol{\theta}) \propto p(\rho)p(\boldsymbol{\alpha}|\boldsymbol{\theta}).$$

Recall that $p(\boldsymbol{\alpha}|\boldsymbol{\theta})$ depends on ρ. The form of the dependence is complex; the expression above gives a nonstandard kernel. However, since the expression, which is proportional to the actual density function, can be evaluated for any $0<\rho<1$, this suggests the use of a Metropolis–Hastings step to generate draws.

A simple random walk is used as the proposal density. See Chib and Greenberg (1996) and Koop and Van Dijk (2000) for related applications of the M–H algorithm. The details are as follows. At iteration i, the candidate ρ^* is chosen as $\rho^* = \rho^{(i-1)} + \upsilon_t$, where υ_t is mean-zero Gaussian with variance determined by the assumed scale. The ratio $R = p(\rho^*|\bullet)/p(\rho^{(i-1)}|\bullet)$ is computed. If $R \leqslant 1$, then $\rho^{(i)}$ is set equal to ρ^* with probability R and to $\rho^{(i-1)}$ with probability $1-R$. Otherwise, if $R>1$, then $\rho^{(i)}$ is set to ρ^* with certainty.

To summarize, in Step 3, the ith draw $\rho^{(i)}$ is produced by first, generating a candidate draw, and second, accepting it or rejecting it (and instead repeating the previous draw) based on a comparison of the kernel $p(\rho|\boldsymbol{\Sigma}_\zeta^{(i)}, \boldsymbol{\Sigma}_\kappa^{(i)}, \boldsymbol{\Sigma}_\varepsilon^{(i)}, \lambda_c^{(i-1)}, \boldsymbol{\alpha}^{(i)}, \mathbf{y})$ at candidate and previous ρ. In this way, the M–H step is embedded in the Gibbs sampler. For each model, prior, and data set, the scale of the proposal density for ρ was calibrated to give an acceptance probability of about 35%. Calibration of the scale was done to improve the efficiency and performance of the simulations.

4. $p(\lambda_c|\boldsymbol{\Sigma}_\zeta^{(i)}, \boldsymbol{\Sigma}_\kappa^{(i)}, \boldsymbol{\Sigma}_\varepsilon^{(i)}, \rho^{(i)}, \boldsymbol{\alpha}^{(i)}, \mathbf{y})$: The conditional posterior is

$$p(\lambda_c|\boldsymbol{\Sigma}_\zeta, \boldsymbol{\Sigma}_\kappa, \boldsymbol{\Sigma}_\varepsilon, \rho, \boldsymbol{\alpha}, \mathbf{y}) \propto p(\lambda_c)p(\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\theta})p(\boldsymbol{\alpha}|\boldsymbol{\theta}) \propto p(\lambda_c)p(\boldsymbol{\alpha}|\boldsymbol{\theta}). \tag{24}$$

The kernel is clearly nonstandard in the frequency parameter: the informative beta priors $p(\lambda_c)$ combine with the factor $p(\boldsymbol{\alpha}|\boldsymbol{\theta})$, which involves an exponential of trigonometric functions of λ_c. However, the quantity proportional to the density function can again be evaluated. A second M–H step is used to compute the successive draws $\lambda_c^{(i)}$ from $p(\lambda_c|\boldsymbol{\Sigma}_\zeta^{(i)}, \boldsymbol{\Sigma}_\kappa^{(i)}, \boldsymbol{\Sigma}_\varepsilon^{(i)}, \rho^{(i)}, \boldsymbol{\alpha}^{(i)}, \mathbf{y})$. The treatment is analogous to Step 3, and the proposal density was calibrated for each application in the same way.

There are various strategies available for estimating the marginal likelihood from the simulation output. Kass and Raftery (1995) and DiCiccio et al. (1997) discuss computational approaches based on the Laplace method. This method relies on a multivariate Gaussian approximating density and has the advantage of simplicity. In our applications of MCMC, the calibration of M–H steps and the discarding of multiple iterations between recorded draws were used to improve the efficiency of the posterior sample. In cases of unusual posterior shapes, for instance, if it were multimodal in a high-density region, then the approximation is likely to break down; see, for instance, Kleijn and Van Dijk (2006). However, the results in our case suggest more or less well-behaved posteriors, at least for the univariate case. Caution may be warranted when making comparisons among multivariate models since in this case the higher-dimensional posteriors involve a more extensive set of relationships between components and parameters.

The MCMC algorithm for the special case of a univariate model has the same form. The primary difference is that the treatment of the variance parameters is now simpler. With each variance matrix now a single variance parameter, in Step 2 of the Gibbs sampler, the joint conditional posterior reduces to a product of three inverted gamma densities.

The use of flat priors on the variance parameters helped avoid distortion of the posterior shape. This was especially important for estimation of the variance of the slope disturbance—this tends to be low when a relatively smooth trend is present in the series. Our experience with simulations showed that the use of an inverted gamma prior density, with shape and scale set to very low values in an attempt to mimic an 'effectively noninformative' prior, could result in distortions in posterior estimates. The marginal posterior of σ_ζ^2 became artificially concentrated in a region near zero, due to the shape of the 'effectively noninformative' IG prior in this region—specifically, its sharp rise in the immediate vicinity of $\sigma_\zeta^2 = 0$. The use of a flat prior enabled a direct picture of the likelihood surface and so gave a clearer indication of the trend variation evident in the data.

For simplicity, the discussion of variance priors has been framed in the univariate context. For multivariate models, a prior on a variance matrix contains information about the covariances between different elements. A correlation structure is implicitly assumed in using the prior. This raises the possibility of more complex effects from the use of 'effectively noninformative' IW priors. In using flat priors for the variance matrices in the applications, the goal is to better preserve the shape of the multivariate likelihood.

The last set of drawings $\{\boldsymbol{\theta}^{(j)}, \boldsymbol{\alpha}^{(j)}\}$, $j = 1, \ldots, J$, used in the final posterior sample was a subset of the $\{\boldsymbol{\theta}^{(i)}, \boldsymbol{\alpha}^{(i)}\}$ produced by the MCMC routine. To reduce correlations between successive posterior draws, multiple iterations were run to produce each draw, and a number of initial iterations were burned. For instance, in the univariate case, 20 iterations were used to produce each draw, and the first 5000 iterations were burned. In the final posterior sample, the autocorrelations for consecutive parameter draws typically fell to near zero after less than 10 lags.

References

Baxter, M., King, R.G., 1999. Measuring business cycles: approximate band-pass filters for economic time series. Review of Economics and Statistics 81, 575–593.

Blanchard, O.J., Fischer, S., 1989. Lectures in Macroeconomics. MIT Press, Cambridge, MA.

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

Chib, S., Greenberg, E., 1996. Markov chain Monte Carlo simulation methods in econometrics. Econometric Theory 12, 409–431.
Cogley, T., 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, 253–278.
De Jong, P., Shephard, N., 1995. The simulation smoother for time series models. Biometrika 82, 339–350.
DiCiccio, T., Kass, R., Raftery, A., Wasserman, L., 1997. Computing Bayes factors by combining simulation and asymptotic approximations. Journal of the American Statistical Association 92, 903–915.
Doornik, J.A., 1999. Ox: An Object-Oriented Matrix Programming Language. Timberlake Consultants Ltd, London.
Durbin, J., Koopman, S.J., 2002. A simple and efficient simulation smoother. Biometrika 89, 603–616.
Frühwirth-Schnatter, S., 1994. Data augmentation and dynamic linear models. Journal of Time Series Analysis 15, 183–202.
Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 3, 515–533.
Harding, D., Pagan, A., 2002. Dissecting the cycle: a methodological investigation. Journal of Monetary Economics 49, 365–381.
Harvey, A.C., 1989. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge.
Harvey, A.C., 2001. Testing in unobserved components models. Journal of Forecasting 20, 1–19.
Harvey, A.C., Jaeger, A., 1993. Detrending, stylised facts and the business cycle. Journal of Applied Econometrics 8, 231–247.
Harvey, A.C., Koopman, S.J., 1997. Multivariate structural time series models. In: Heij, C., et al. (Eds.), System Dynamics in Economic and Financial Models. Wiley, Chichester.
Harvey, A.C., Trimbur, T.M., 2003. General model-based filters for extracting trends and cycles in economic time series. Review of Economics and Statistics 85, 244–255.
Harvey, A.C., Trimbur, T.M., Van Dijk, H.K., 2005. Bayes estimates of the cyclical component in twentieth century US gross domestic product. In: Proceedings of the 4th Eurostat and DGECFIN Colloquium on Modern Tools for Business Cycle Analysis, Eurostat.
Huerta, G., West, M., 1999a. Priors and component structures in autoregressive time series models. Journal of the Royal Statistical Society, Series B 61, 881–899.
Huerta, G., West, M., 1999b. Bayesian inference on periodicities and component spectral structure in time series. Journal of Time Series Analysis 20, 401–416.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. Journal of the American Statistical Association 24, 773–795.
King, R., Plosser, C., Stock, J., Watson, M., 1991. Stochastic trends and economic fluctuations. American Economic Review 81, 819–840.
Kleijn, R., Van Dijk, H.K., 2006. Bayes model averaging of cyclical decompositions in economic time series. Journal of Applied Econometrics 21, 191–212.
Koop, G., Van Dijk, H.K., 2000. Testing for integration using evolving trend and seasonals models: a Bayesian approach. Journal of Econometrics 97, 261–291.
Koopman, S.J., Shephard, N., Doornik, J., 1999. Statistical algorithms for models in state space using SsfPack 2.2. Econometrics Journal 2, 113–166.
Koopman, S.J., Harvey, A.C., Doornik, J.A., Shephard, N., 2000. STAMP 6.0 Structural Time Series Analysis Modeller and Predictor. Timberlake Consultants Ltd, London.
Kuttner, K., 1994. Estimating potential output as a latent variable. Journal of Business and Economic Statistics 12, 361–368.
Magnus, J., Neudecker, H., 1999. Matrix Differential Calculus with Applications in Statistics and Econometrics. Wiley, Chichester.
Murray, C.J., 2003. Cyclical properties of Baxter–King filtered time series. Review of Economics and Statistics 85, 471–476.
Orphanides, A., Van Norden, S., 2002. The unreliability of output gap estimates in real-time. Review of Economics and Statistics 84, 569–583.
Planas, C., Rossi, A., 2004. Can inflation data improve the real-time reliability of output gap estimates? Journal of Applied Econometrics 19, 121–133.
Poirier, D., 1995. Intermediate Statistics and Econometrics. MIT Press, Cambridge, MA.
Stoffer, D., Wall, K., 2004. Resampling in state space models. In: Harvey, A.C., Koopman, S.J., Shephard, N. (Eds.), State Space and Unobserved Component Models. Cambridge University Press, Cambridge, pp. 171–202.

Trimbur, T.M., 2006. Properties of higher order stochastic cycles. Journal of Time Series Analysis 27, 1–17.
Zellner, A., 1971. An Introduction to Bayesian Inference in Econometrics. Wiley, New York.
Zellner, A., Hong, C., Gulati, G., 1990. Turning points in economic time series, loss structures, and Bayesian forecasting. In: Geisser, S., et al. (Eds.), Bayesian and Likelihood Methods in Statistics and Econometrics. North-Holland, Elsevier Science Publishers, Amsterdam.