Structural Time Series Models for Business Cycle Analysis

Proietti, Tommaso

SEFEMEQ, Faculty of Economics, University of Rome "Tor Vergata"

20 January 2008

Structural Time Series Models for Business Cycle Analysis *

Tommaso Proietti[†]
S.E.F. e ME.Q., University of Rome "Tor Vergata".

Abstract

The chapter deals with parametric models for the measurement of the business cycle in economic time series. It presents univariate methods based on parametric trend–cycle decompositions and multivariate models featuring a Phillips type relationship between the output gap and inflation and the estimation of the gap using mixed frequency data. We finally address the issue of assessing the accuracy of the output gap estimates.

Keywords: State Space Models. Kalman Filter and Smoother. Bayesian Estimation.

Contents

1 Introduction **2**

2 Univariate Methods **3**

2.1 The random walk plus noise model . . . 4
2.2 The local linear model and the Leser-HP filter . . . 5
2.3 Higher order trends and low-pass filters . . . 7
2.4 The cyclical component . . . 8
2.5 Models with correlated components . . . 10
2.6 Model–based band-pass filters . . . 14
2.7 Applications of model-based filtering: band-pass cycles and the estimation of recession probabilities . . . 16
2.8 Ad-hoc filtering and the Slutsky-Yule effect . . . 19

3 Multivariate Models **21**

3.1 Bivariate models of real output and inflation . . . 21
3.2 A bivariate quarterly model of output and inflation for the U.S. . . . 22
3.2.1 Maximum likelihood estimation . . . 23
3.2.2 Bayesian estimation . . . 24
3.3 Multivariate extensions . . . 28
3.4 A multivariate model with mixed frequency data . . . 31

4 The Reliability of the Output Gap Measurement **32**

4.1 Validity . . . 34
4.2 Precision . . . 34

*To appear in the *Handbook of Econometrics: Vol. 2, Applied Econometrics*, Part 3.4., ed. T. Mills and K. Patterson, Palgrave, London, forthcoming, 2008.

[†]Address for Correspondence: Via Columbia 2, 000133 Rome, Italy. Tel. +39 06 7259 5941. E-mail: tommaso.proietti@uniroma2.it. The author wishes to thank Terence Mills and Alberto Musso for useful suggestions.

A Linear filters 36

B The Wiener-Kolmogorov filter 36

C State space models and methods 37
C.1 The augmented Kalman filter . . . 38
C.2 Real time (updated) estimates . . . 38
C.3 Smoothing . . . 38
C.4 The simulation smoother . . . 39

References 40

1 Introduction

The term structural time series refers to a class of parametric models that are specified directly in terms of unobserved components which capture essential features of the series, such as trends, cycles and seasonality. The approach is amenable to the analysis of macroeconomic time series, where latent variables such as trends and cycles, and more specialised notions, such as the output gap, core inflation and the natural rate of unemployment, need to be measured.

One of the key issues economists have faced in characterising the dynamic behaviour of macroeconomic variables, such as output, unemployment and inflation, is separating trends from cycles. The decomposition of economic time series has a long tradition, dating back to the 19th century; see the first chapter of Mills (2003) for an historical perspective. Along with providing a description of the salient features of a series, the distinction of what is permanent and what is transitory in economic dynamics bears relevant implications for monetary and fiscal policy. The underlying idea is that trends and cycles can be ascribed to different economic mechanisms and an understanding of their determinants helps to define policy targets and instruments.

This chapter focusses on structural time series modelling for business cycle analysis and, in particular, for output gap measurement. The *output gap* is the deviation of the economy's realised output from its potential. Potential output is defined as the non-inflationary level of output, i.e., as the level that can be attained using the available technology and productive factors at a stable inflation rate. The gap measures the presence and the extent of real disequilibria and constitutes an indicator of inflationary pressure in the short run: a positive output gap testifies to excess demand and a negative output gap expresses excess supply.

The output gap plays a central role in the transmission mechanism of monetary policy, since short term interest rates influence aggregate demand and the latter affects inflation via a Phillips curve relationship. The Phillips curve establishes a trade-off between output and inflation over the short run, and provides the rationale for using the short run component in output as an indication of demand-driven inflationary pressure. For instance, the Taylor rule (Taylor, 1999) explicitly links the central bank's policy to the output gap. On the other hand, the growth rate of potential output is a reference value for broad money growth. Other important uses of the output gap are in fiscal analysis, where it is employed to assess the impact of cyclical factors on budget deficits, and in the adjustment of exchange rates. The output gap is also related to cyclical unemployment, which is the deviation of unemployment from its trend, known as the non-accelerating-inflation rate of unemployment (NAIRU).

The signal extraction problems relating to latent variables, such as the output gap, core inflation and the NAIRU, can be consistently formulated within a model based framework and, in particular, within the class of unobserved components time series models, formalising the fundamental economic relationships with observable macroeconomic aggregates.

The chapter is divided into three main parts: the first (section 2) deals with univariate methods for cycle measurement. One approach is to formalise a model of economic fluctuations such that the different components are driven by specific shocks, that are propagated via a dynamic transmission mechanism. We start introducing the traditional trend-cycle structural decomposition, discussing the parametric representation of both components (sections 2.1-2.4), and the correlation between the trend and cycle disturbances (section 2.5). Another approach is to consider the cycle as the band-pass component of output, i.e. as those economic fluctuations which have a periodicity greater then a year and smaller than say eight years. We review the relationship between popular signal extraction filters such as the Hodrick-Prescott filter and the Baxter and King filter, and the model-based Wiener-Kolmogorov filter. Particular attention is devoted to the implementation of band-pass filtering in a model-based framework (section 2.6). The advantages of this strategy are twofold: the components can be computed also in real time using standard principles of optimal signal extraction, and thus efficient algorithms, such as the Kalman filter and smoother, can be applied. Secondly, the reliability of the estimated components can be thoroughly assessed.

The second part, starting with section 3, deals with multivariate models for the measurement of the output gap. The above definition of the output gap as an indicator of inflationary pressures suggests that the minimal most basic measurement framework is a bivariate model for output and inflation. After reviewing the work done in this area (section 3.1) we illustrate the estimation a bivariate model for the U.S. economy, under both the classical and the Bayesian approach and incorporating the feature known as "great moderation" of the volatility of economic fluctuations (section 3.2). In section 3.3 we review the multivariate extensions of the basic bivariate model and we conclude this part with an application which serves to illustrate the flexibility of the state space methodology in accommodating data features such as missing data, nonlinearities and temporal aggregation. In particular, section 3.4 presents the results of fitting a four variate monthly time series model for the U.S. economy with mixed frequency data, as gross domestic product (GDP) is available only quarterly, whereas industrial production, the unemployment rate and inflation are monthly. The model incorporates the temporal aggregation constraints (which are nonlinear since the model is formulated in terms of the logarithm of the variables) and produces as a byproduct monthly estimates of GDP, along with their reliability, that are consistent with the quarterly observed values.

The third part 4 deals with the reliability of the output gap estimates. The assessment of the quality of the latter is crucial for the decision maker. We discuss the various sources of uncertainty (model selection, parameter estimation, data revision, estimation of unboserved components, statistical revision), and discuss ways of dealing with them using the state space methodology.

One of the objectives of this chapter is to provide an overview of the main state space methods and to illustrate their application and scope. The description of the algorithms is relegated to an appendix and we refer to Harvey (1989), West and Harrison (1997), Kitagawa and Gersch (1996), Durbin and Koopman (2000), and the selection of readings in Harvey and Proietti (2005), for a thorough presentation of the main ideas and methodological aspects concerning state space methods and unobserved components models. For the class of state space models with Markov switching, see Kim and Nelson (1999b), Frühwirth-Schnatter (2006) and Cappé *et al.* (2005). An essential and up to date monograph on modelling trends and cycles in economics is Mills (2003).

2 Univariate Methods

In univariate analysis, the output gap can be identified as the stationary or transitory component in a measure of aggregate economic activity, such as GDP. Estimating the output gap thus amounts to *detrending* the series; a large literature has been devoted to this very controversial issue (see, for instance, Canova, 1998, and Mills, 2003).

We shall confine our attention to the additive decomposition (after a logarithmic transformation) of real output, y_t, into potential output, μ_t, and the output gap, ψ_t: $y_t = \mu_t + \psi_t$. This basic representation is readily extended to handle a seasonal component and other calendar components such as those associated with trading days and moving festivals, which for certain output series, e.g. industrial production, play a relevant role.

In the structural approach a parametric representation for the components is needed; furthermore, the specification of the model is completed by assumptions on the covariance among the various components. The first identifying restriction that will be adopted throughout is that μ_t is fully responsible for the nonstationary behaviour of the series, whereas ψ_t is a transitory component.

2.1 The random walk plus noise model

The random walk plus noise (RWpN) model provides the most basic trend-cycle decomposition of output, such that the trend is a random walk process, with normal and independently distributed (NID) increments, and the cycle is a pure white noise (WN) component. The structural specification is the following:

$$\begin{array}{rcll} y_t & = & \mu_t + \psi_t, \qquad t = 1, \ldots, n, & \psi_t \sim \mathrm{NID}(0, \sigma^2_\psi), \\ \mu_t & = & \mu_{t-1} + \beta + \eta_t, & \eta_t \sim \mathrm{NID}(0, \sigma^2_\eta). \end{array} \tag{1}$$

When the drift is absent, i.e., when $\beta = 0$, the model is also known as the *local level model*, see Harvey (1989). We assume throughout that $\mathrm{E}(\eta_t \psi_{t-j}) = 0$ for all t and j, so the two components are orthogonal.

If $\sigma^2_\eta = 0$, μ_t is a deterministic linear trend. The one-sided Lagrange Multiplier test of the null hypothesis $H_0 : \sigma^2_\eta = 0$, against the alternative $H_1 : \sigma^2_\eta > 0$, is known as a stationarity test and is discussed in Nyblom and Mäkeläinen (1983). The nonparametric extension to the case when ψ_t is any indeterministic stationary process is provided by Kwiatkowski *et al.* (1992). See also Harvey (2001) for a review and extensions.

The reduced-form representation of (1) is an integrated moving average model of orders (1,1), or IMA(1,1): $\Delta y_t = \beta + \xi_t + \theta \xi_{t-1}, \xi_t \sim \mathrm{NID}(0, \sigma^2)$, where $\Delta y_t = y_t - y_{t-1}$. The difference operator can be defined in terms of the lag operator L, such that $L^d y_t = y_{t-d}$, for an integer d, as $\Delta = (1 - L)$.

The moving average (MA) parameter is subject to the restriction $-1 \leq \theta \leq 0$. Equating the autocovariance generating functions of Δy_t implied by the IMA(1,1) and by the structural representation (1), it is possible to establish that $\sigma^2_\eta = (1+\theta)^2 \sigma^2$ and $\sigma^2_\psi = -\theta \sigma^2$. Hence, it is required that $\theta \leq 0$, and thus persistence, $(1 + \theta)$, cannot be greater than unity. The variance ratio $\lambda = \sigma^2_\psi / \sigma^2_\eta$ depends uniquely on θ, as $\lambda = -\theta/(1+\theta)^2$. The ratio provides a measure of relative smoothness of the trend: if λ is large, then the trend varies little with respect to the noise component, and thus it can be regarded as "smooth".

The RWpN model has a long tradition and a well-established role in the analysis of economic time series, since it provides the model-based interpretation for the popular forecasting technique known as *exponential smoothing*, which is widely used in applied economic forecasting and fares remarkably well in forecast competitions; see Muth (1960) and the comprehensive reviews by Gardner (1985, 2006).

Assuming a doubly infinite sample, the one-step-ahead predictions, $\tilde{\mu}_{t+1|t}$, and the filtered and smoothed estimates of the trend component, denoted $\tilde{\mu}_{t|\infty}$, are given, respectively, by:

$$\tilde{\mu}_{t+1|t} = \tilde{\mu}_{t|t} = (1+\theta) \sum_{j=0}^{\infty} (-\theta)^j y_{t-j}, \quad \tilde{\mu}_{t|\infty} = \frac{1+\theta}{1-\theta} \sum_{j=-\infty}^{\infty} (-\theta)^{|j|} y_{t-j}.$$

Here, $\tilde{\mu}_{t+1|t}$ denotes the expectation of μ_{t+1} based on the information available at time t, whereas $\tilde{\mu}_{t|\infty}$ is the expectation based on all of the information in the doubly infinite data set. The filter $\mathrm{w}(L) = (1+\theta)(1+\theta L)^{-1} = (1+\theta)\sum_{j=0}^{\infty}(-\theta)^j L^j$ is known as a one–sided exponentially weighted moving average (EWMA). These expressions follow from applying the Wiener–Kolmogorov prediction and signal extraction formulae; see appendix B. In terms of the structural form parameters, $\tilde{\mu}_{t|\infty} = \frac{\sigma^2_\eta}{\sigma^2_\eta + \sigma^2_\psi |1-L|^2} y_t$, where $|1-L|^2 = (1-L)(1-L^{-1})$. The filter

$$\mathrm{w}_\mu(L) = \frac{\sigma^2_\eta}{\sigma^2_\eta + \sigma^2_\psi |1-L|^2} = \frac{1+\theta}{1-\theta} \sum_{j=-\infty}^{\infty} (-\theta)^{|j|} L^j,$$

is known as a two sided EWMA filter. In finite samples, the computations are performed by the Kalman filter and smoother (see appendix C).

The parameter θ (or, equivalently, λ) is essential in determining the weights that are attached to the observations for signal extraction and prediction. When $\theta = 0$, y_t is a pure random walk, and then the current observation provides the best estimate of the trend: $\tilde{\mu}_{t+1|t} = \tilde{\mu}_{t|t} = \tilde{\mu}_{t|\infty} = y_t$. When $\theta = -1$, the trend estimate, which is as smooth as possible, is a straight line passing through the observations.

The RWpN model provides a stripped to the bone separation of the transitory and the permanent dynamics that depends on a single smoothness parameter, which determines the weights that are assigned to the available observations for forecasting and trend estimation. Its use as a misspecified model of economic fluctuations for out-of-sample forecasting, using multistep (or adaptive) estimation, rather than maximum likelihood estimation, has been considered in the seminal paper by Cox (1961), and by Tiao and Xu (1993). Proietti (2005) discusses multistep estimation of the RWpN model for the extraction of trends and cycles.

2.2 The local linear model and the Leser-HP filter

In the local linear trend model (LLTM) the trend μ_t is an integrated random walk:

$$\begin{array}{rclllll} y_t & = & \mu_t & + & \psi_t, & & \psi_t \sim \mathrm{NID}(0, \sigma^2_\psi), \quad t = 1, 2, \ldots, n, \\ \mu_t & = & \mu_{t-1} & + & \beta_{t-1} & + \; \eta_t, & \eta_t \sim \mathrm{NID}(0, \sigma^2_\eta), \\ \beta_t & = & & & \beta_{t-1} & + \; \zeta_t, & \zeta_t \sim \mathrm{NID}(0, \sigma^2_\zeta). \end{array} \tag{2}$$

It is assumed that the ψ_t, η_t and ζ_t are mutually and serially uncorrelated. For $\sigma^2_\zeta = 0$ the trend reduces to a random walk with constant drift, whereas for $\sigma^2_\eta = 0$ the trend is an integrated random walk ($\Delta^2 \mu_t = \zeta_{t-1}$).

The above representation encompasses a deterministic linear trend, arising when both σ^2_η and σ^2_ζ are zero. Secondly, it is consistent with the notion that the real time estimate of the trend is coincident with the value of the eventual forecast function at the same time (see section 2.5 on the Beveridge-Nelson decomposition).

The LLTM is the model for which the Leser filter is optimal (see Leser, 1961). The latter is derived as the minimiser, with respect to $\mu_t, t = 1, \ldots, n$, of the penalised least squares criterion:

$$PLS \quad = \quad \textstyle\sum_{t=1}^{n} (y_t - \mu_t)^2 + \lambda \sum_{t=3}^{n} (\Delta^2 \mu_t)^2.$$

The parameter λ governs the trade-off between fidelity and it is referred to as the *smoothness* or *roughness penalty* parameter. The first addend of PLS measures the goodness of fit, whereas the second penalises the departure from zero of the variance of the second differences (i.e. a measure of roughness). In matrix notation, if $\mathbf{y} = (y_1, \ldots, y_n)$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$, and $\mathbf{D} = \{d_{ij}\}$ is the $n \times n$ matrix corresponding to a first difference filter, with $d_{ii} = 1$, $d_{i,i-1} = -1$ and zero otherwise, so that $\mathbf{D}\boldsymbol{\mu} = (\mu_2 - \mu_1, \ldots, \mu_n - \mu_{n-1})'$, we can write the criterion function as

$PLS = (\mathbf{y}-\boldsymbol{\mu})'(\mathbf{y}-\boldsymbol{\mu})+\lambda\boldsymbol{\mu}'\mathbf{D}^{2'}\mathbf{D}^2\boldsymbol{\mu}$. Differentiating with respect to $\boldsymbol{\mu}$, the first order conditions yield: $\tilde{\boldsymbol{\mu}} = (\mathbf{I}_n+\lambda\mathbf{D}^{2'}\mathbf{D}^2)^{-1}\mathbf{y}$. The rows of the matrix $(\mathbf{I}_n+\lambda\mathbf{D}^{2'}\mathbf{D}^2)^{-1}$ contain the filter weights for estimating the trend at a particular point in time. The solution arising for $\lambda = 1600$ is widely popular in the analysis of quarterly macroeconomic time series as the Hodrick-Prescott filter (HP henceforth, see Hodrick and Prescott, 1997); the choice of the smoothness parameter for yearly and monthly time series is discussed in Ravn and Uhlig (2002) and Maravall and del Rio (2007).

We now show that the Leser filter is the optimal signal extraction filter for the LLTM (2) with $\sigma^2_\eta = 0$ and $\lambda = \sigma^2_\psi/\sigma^2_\zeta$. In fact, apart from an additive term which does not depend on μ, PLS is proportional to $\ln f(\mathbf{y},\boldsymbol{\mu}) = \ln f(\mathbf{y}|\boldsymbol{\mu}) + \ln f(\boldsymbol{\mu})$, where $f(\mathbf{y},\boldsymbol{\mu})$, $f(\mathbf{y}|\boldsymbol{\mu})$ denote, respectively, the gaussian joint density of the random vectors $\mathbf{y}$ and $\boldsymbol{\mu}$, and the conditional density of $\mathbf{y}$ given $\boldsymbol{\mu}$, whereas $f(\boldsymbol{\mu})$ is the joint density of $\boldsymbol{\mu}_t, t = 1,\dots,n$. Now, $\ln f(\mathbf{y}|\boldsymbol{\mu})$ depends on $\boldsymbol{\mu}$ only via $(1/\sigma^2_\psi)\sum_{t=1}^n(y_t-\mu_t)^2$, whereas $\ln f(\boldsymbol{\mu}) = \ln f(\mu_3,\dots,\mu_n|\mu_1,\mu_2) + \ln f(\mu_1,\mu_2)$. The first term depends on $\mu_t, t > 2$, only via $(1/\sigma^2_\zeta)\sum_{t=3}^n(\Delta^2\mu_t)^2$. The contribution of the initial conditions vanishes under fixed initial conditions or diffuse initial conditions[1]. In conclusion, $\tilde{\boldsymbol{\mu}}$ maximises with respect to $\boldsymbol{\mu}$ the joint log-density $\ln f(\mathbf{y},\boldsymbol{\mu})$ and thus the posterior log-density $\ln f(\boldsymbol{\mu}|\mathbf{y}) = \ln f(\mathbf{y},\boldsymbol{\mu}) - \ln f(\mathbf{y})$. A consequence of this result is that the components can be efficiently computed using the Kalman filter and smoother (see appendix C). The latter computes the mean of the conditional distribution $\boldsymbol{\mu}|\mathbf{y}$. As this distribution is Gaussian, the posterior mean is equal to the posterior mode. Hence, the smoother computes the mode of $f(\boldsymbol{\mu}|\mathbf{y})$, which is also the minimiser of the PLS criterion.

The equivalence $\lambda = \sigma^2_\psi/\sigma^2_\zeta$ makes clear that the roughness penalty measures the variability of the cyclical (noise) component relative to that of the trend disturbance, and regulates the smoothness of the long-term component. As σ^2_ζ approaches zero, λ tends to infinity, and the limiting representation of the trend is a straight line. The Leser-HP detrended or cyclical component is the smoothed estimate of the component ψ_t in (2) and, although the maintained representation for the deviations from the trend is a WN component, the filter has been one of the most widely employed tools in macroeconomics to extract a measure of the business cycle. For the U.S. GDP series (logarithms) this component is plotted in the top right hand panel of figure 4.

In terms of the reduced form of model (2), the IMA(2,2) model $\Delta^2 y_t = (1+\theta_1 L+\theta_2 L^2)\xi_t$, $\xi_t \sim \mathrm{NID}(0,\sigma^2)$, it can be shown that the restriction $\sigma^2_\eta = 0$ implies $[(1+\theta_2)\theta_2]/(1-\theta_2)^2 = \lambda$ and $\theta_1 = -4\theta_2/(1+\theta_2)$. Therefore, for $\lambda = 1600$, we have $\theta_1 = -1.778$ and $\theta_2 = 0.799$, so that $\theta(1) = 1+\theta_1+\theta_2 = 0.021$ and the MA polynomial is close to noninvertibility at the zero frequency.

The theoretical properties of the Leser-HP filter are better understood by assuming the availability of a doubly infinite sample, $y_{t+j}, j = -\infty,\dots,\infty$. In such a setting, the Wiener-Kolmogorov filter (see Whittle, 1963, and appendix B) provides the minimum mean square linear estimator (MMSLE) of the trend, $\tilde{\mu}_{t|\infty} = \mathrm{w}_\mu(L)y_t$, where

$$\mathrm{w}_\mu(L) = \frac{\sigma^2_\zeta}{\sigma^2_\zeta + |1-L|^4\sigma^2_\psi} = \frac{1}{1+\lambda|1-L|^4} \tag{3}$$

The frequency response function of the trend filter (see appendix A) is:

$$\mathrm{w}_\mu(e^{-\imath\omega}) = \frac{1}{1+4\lambda(1-\cos\omega)^2}, \quad \omega \in [0,\pi];$$

notice that this is 1 at the zero frequency and decreases monotonically to zero as ω approaches π. This behaviour enforces the interpretation of (3) as a low-pass filter, and the corresponding

[1]Assuming $\boldsymbol{\mu}_* = (\mu_1,\mu_2)' \sim \mathrm{N}(\mathbf{0},\Sigma_\mu)$, and that the process μ_t has started in the indefinite past, $\boldsymbol{\Sigma}_\mu^{-1} \to \mathbf{0}$, and thus the quadratic form $\boldsymbol{\mu}_*'\boldsymbol{\Sigma}_\mu^{-1}\boldsymbol{\mu}_*$ converges to zero.

detrending filter, $1 - \mathrm{w}_\mu(L)$, is the high-pass filter derived from it. We shall return to this issue in the next section.

2.3 Higher order trends and low-pass filters

A *low-pass* filter is a filter that passes low frequency fluctuations and reduces the amplitude of fluctuations with frequencies higher than a cutoff frequency ω_c (see, e.g., Percival and Walden, 1993). The frequency response function of an ideal low-pass filter takes the following form: for $\omega \in [0, \pi]$,

$$\mathrm{w}_{lp}(\omega) = \begin{cases} 1 & \text{if } \omega \leq \omega_c \\ 0 & \text{if } \omega_c < \omega \leq \pi \end{cases}$$

The notion of a high-pass filter is complementary, its frequency response function being $\mathrm{w}_{hp}(\omega) = 1 - \mathrm{w}_{lp}(\omega)$. The coefficients of the ideal low-pass filter are provided by the inverse Fourier transform of $\mathrm{w}_{lp}(\omega)$:

$$\mathrm{w}_{lp}(L) = \frac{\omega_c}{\pi} + \sum_{j=1}^{\infty} \frac{\sin(\omega_c j)}{\pi j}(L^j + L^{-j}).$$

A band-pass filter is a filter that passes fluctuations within a certain frequency range and attenuates those outside that range. Given lower and upper cutoff frequencies, $\omega_{1c} < \omega_{2c}$ in $(0, \pi)$, the ideal frequency response function is unity in the interval $[\omega_{1c}, \omega_{2c}]$ and zero outside. The notion of a band–pass filter is relevant to business cycle measurement: the traditional definition, ascribed to Burns and Mitchell (1946), considers all the fluctuations with a specified range of periodicities, namely those ranging from one and a half to eight years. Thus, if s is the number of observations in a year, the fluctuations with periodicity between $1.5s$ and $8s$ are included. Baxter and King (1999, BK henceforth) argue that the ideal filter for cycle measurement is a band-pass filter. Now, given the two business cycle frequencies, $\omega_{c1} = 2\pi/(8s)$ and $\omega_{c2} = 2\pi/(1.5s)$, the band-pass filter is

$$\mathrm{w}_{bp}(L) = \frac{\omega_{c2} - \omega_{c1}}{\pi} + \sum_{j=1}^{\infty} \frac{\sin(\omega_{c2} j) - \sin(\omega_{c1} j)}{\pi j}(L^j + L^{-j}). \tag{4}$$

Notice that $w_{bp}(L)$ is the contrast between the two low–pass filters with cutoff frequencies ω_{c2} and ω_{c1}. The frequency response function of the ideal business cycle band-pass filter for quarterly observations ($s = 4$), which is equivalent to the gain function (see Appendix A), is plotted in figure 3.

The ideal band-pass filter exists and is unique, but as it entails an infinite number of leads and lags, an approximation is required in practical applications. BK show that the K-terms approximation to the ideal filter (4), which is optimal in the sense of minimising the integrated mean square approximation error, is obtained from (4) by truncating the lag distribution at a finite integer K. They propose using a three years window, i.e., $K = 3s$, as a valid rule of thumb for macroeconomic time series. They also constrain the weights to sum to zero, so that the resulting approximation is a detrending filter: denoting the truncated filter $\mathrm{w}_{bp,K}(L) = \mathrm{w}_0 + \sum_1^K \mathrm{w}_j(L^j + L^{-j})$, the weights of the adjusted filter will be $\mathrm{w}_j - \mathrm{w}_{bp,K}(1)/(2K + 1)$. The gain of the resulting filter is displayed in figure 3 (henceforth we shall refer to it as the BK filter). The ripples result from the truncation of the ideal filter and are referred to as the Gibbs phenomenon (see Percival and Walden, 1993, p. 177). BK do not entertain the problem of estimating the cycle at the extremes of the available sample; as a result the estimates for the first and last three years are unavailable. Christiano and Fitzgerald (2003) provide the optimal finite-sample approximations for the band pass filter, including the real time filter, using a model based approach.

Within the class of parametric structural models, an important category of low–pass filters emerges from the application of Wiener-Kolmogorov optimal signal extraction theory to the following model:

$$\begin{array}{lcll} y_t & = & \mu_t + \psi_t, & t = 1, 2, \ldots, n, \\ \Delta^m \mu_t & = & (1+L)^r \zeta_t, \quad \zeta_t \sim \text{NID}(0, \sigma^2_\zeta), & \\ \psi_t & \sim & \text{NID}(0, \lambda\sigma^2_\zeta), \quad \text{E}(\zeta_t, \psi_{t-j}) = 0, \forall j, & \end{array} \tag{5}$$

where μ_t is the signal or trend component, and ψ_t is the noise.

Assuming a doubly infinite sample, the minimum mean square estimators of the components (see Appendix B) are, respectively, $\tilde{\mu}_t = \text{w}_\mu(L)y_t$ and $\tilde{\psi}_t = y_t - \tilde{\mu}_t = [1 - \text{w}_\mu(L)]y_t$, where

$$\text{w}_\mu(L) = \frac{|1+L|^{2r}}{|1+L|^{2r} + \lambda|1-L|^{2m}}. \tag{6}$$

The expression (6) defines a class of filters which depends on the order of integration of the trend (m, which regulates its flexibility), on the number of unit poles at the Nyquist frequency r, which *ceteris paribus* regulates the smoothness of $\Delta^m \mu_t$, and λ, which measures the relative variance of the noise component.

The Leser-HP filter arises for $m = 2, r = 0, \lambda = 1600$ (quarterly data). The two-sided EWMA filter arises for $m = 1, r = 0$. The filters arising for $m = r$ are Butterworth filters of the tangent version (see, e.g., Gómez, 2001). The analytical expression of the gain is:

$$\text{w}_\mu(\omega) = \left\{1 + \left[\frac{\tan(\omega/2)}{\tan(\omega_c/2)}\right]^{2m}\right\}^{-1},$$

and depends solely on m and ω_c. As $m \to \infty$ the gain converges to the frequency response function of the ideal low–pass filter.

The previous discussion enforces the interpretation of the trend filter $\text{w}_\mu(L)$ as a low–pass filter. Its cut-off frequency depends on the triple (m, r, λ). Frequency domain arguments can be advocated for designing these parameters so as to select the fluctuations that lie in a pre-determined periodicity range. In particular, let us consider the Fourier transform of the trend filter (6), $\text{w}_\mu(\omega) = \text{w}_\mu(e^{-\imath\omega}), \omega \in [0, \pi]$, which also expresses the gain of the filter. The latter is monotonically decreasing with λ; it takes the value 1 at the zero frequency and, if $r > 0$, it is zero at the Nyquist frequency. The trend filter will preserve to a great extent those fluctuations at frequencies for which the gain is greater than 1/2 and reduce to a given extent those for which the gain is below 1/2. This simple argument justifies the definition of a low–pass filter with cutoff frequency ω_c if the gain halves at that frequency; see Gomez (2001, sec. 1). Usually the investigator sets the cut-off frequency to a particular value, e.g. $\omega_c = 2\pi/(8s)$ and chooses the values of m and r (e.g., $m = 2, r = 0$ for the Leser- HP filter). Solving the equation $\text{w}_\mu(\omega_c) = 1/2$, the parameter λ can be obtained in terms of the cut-off frequency and the orders m and r:

$$\lambda = 2^{r-m}\left[\frac{(1+\cos\omega_c)^r}{(1-\cos\omega_c)^m}\right]. \tag{7}$$

2.4 The cyclical component

In the previous section we considered some of the most popular decompositions of a time series into a trend and pure white noise component. Hence, the previous models are misspecified. In the analysis of economic time series it is more interesting to entertain a trend-cycle decomposition, such that the trend is due to the accumulation of supply shocks that are permanent, whereas the cycle is ascribed to nominal or demand shocks that are propagated by a stable transmission mechanism. Clark (1987) and Harvey and Jäger (1993), for instance, replace the irregular component by a stationary stochastic cycle, which is parameterised as an AR(2) or an ARMA(2,1) process, such that the roots of the AR polynomial are a pair of complex conjugates. The model for the cycle is a stationary process capable of reproducing widely acknowledged stylised facts, such as the presence of strong autocorrelation, determining the recurrence and alternation of phases, and the dampening of fluctuations, or zero long run persistence.

In particular, the model adopted by Clark (1987) is:

$$\psi_t = \phi_1 \psi_{t-1} + \phi_2 \psi_{t-2} + \kappa_t, \quad \kappa_t \sim \mathrm{NID}(0, \sigma^2_\kappa),$$

where κ_t is independent of the trend disturbances. Harvey (1989) and Harvey and Jäger (1993) use a different representation:

$$\begin{bmatrix} \psi_t \\ \psi^*_t \end{bmatrix} = \rho \begin{bmatrix} \cos\varpi & \sin\varpi \\ -\sin\varpi & \cos\varpi \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi^*_{t-1} \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa^*_t \end{bmatrix}, \tag{8}$$

where $\kappa_t \sim \mathrm{NID}(0, \sigma^2_\kappa)$ and $\kappa^*_t \sim \mathrm{NID}(0, \sigma^2_\kappa)$ are mutually independent and independent of the trend disturbance, $\varpi \in [0, \pi]$ is the frequency of the cycle and $\rho \in [0, 1)$ is the damping factor. The reduced form of (8) is the ARMA(2,1) process:

$$(1 - 2\rho\cos\varpi L + \rho^2 L^2)\psi_t = (1 - \rho\cos\varpi L)\kappa_t + \rho\sin\varpi\kappa^*_{t-1}.$$

When ρ is strictly less than one the cycle is stationary with $\mathrm{E}(\psi_t) = 0$ and $\sigma^2_\psi = \mathrm{Var}(\psi_t) = \sigma^2_\kappa/(1-\rho^2)$; the autocorrelation at lag j is $\rho^j \cos\varpi j$. For $\varpi \in (0, \pi)$ the roots of the AR polynomial are a pair of complex conjugates with modulus ρ^{-1} and phase ϖ; correspondingly, the spectral density displays a peak around ϖ.

Harvey and Trimbur (2002) further extend the model specification, by proposing a general class of model based filters for extracting trend and cycles in macroeconomic time series, showing that the design of low-pass and band-pass filters can be considered as a signal extraction problem in an unobserved components framework. In particular, they consider the decomposition $y_t = \mu_{mt} + \psi_{kt} + \epsilon_t$, where $\epsilon_t \sim \mathrm{NID}(0, \sigma^2_\epsilon)$. The trend is specified as an m-th order stochastic trend:

$$\begin{array}{lclcll} \mu_{1t} & = & \mu_{1,t-1} & + & \zeta_t & \\ \mu_{it} & = & \mu_{i,t-1} & + & \mu_{i-1,t}, & \qquad i = 2, \ldots, m \end{array} \tag{9}$$

This is the recursive definition of an $m-1$-fold integrated random walk, such that $\Delta^m \mu_{mt} = \zeta_t$. The component ψ_{kt} is a k-th order stochastic cycle, defined as:

$$\begin{bmatrix} \psi_{1t} \\ \psi^*_{1t} \end{bmatrix} = \rho \begin{bmatrix} \cos\varpi & \sin\varpi \\ -\sin\varpi & \cos\varpi \end{bmatrix} \begin{bmatrix} \psi_{1,t-1} \\ \psi^*_{1,t-1} \end{bmatrix} + \begin{bmatrix} \kappa_t \\ 0 \end{bmatrix},$$

$$\begin{bmatrix} \psi_{it} \\ \psi^*_{it} \end{bmatrix} = \rho \begin{bmatrix} \cos\varpi & \sin\varpi \\ -\sin\varpi & \cos\varpi \end{bmatrix} \begin{bmatrix} \psi_{i,t-1} \\ \psi^*_{i,t-1} \end{bmatrix} + \begin{bmatrix} \psi_{i-1,t} \\ 0 \end{bmatrix}, \tag{10}$$

The reduced form representation for the cycle is:

$$(1 - 2\rho\cos\varpi L + \rho^2 L^2)^k \psi_{kt} = (1 - \rho\cos\varpi L)^k \kappa_t.$$

Harvey and Trimbur show that, as m and k increase, the optimal estimators of the trend and the cycle approach the ideal low-pass and band-pass filter, respectively.

2.5 Models with correlated components

Morley, Nelson and Zivot (2003, MNZ henceforth) consider the following unobserved components model for U.S. quarterly GDP:

$$
\begin{aligned}
y_t &= \mu_t + \psi_t & t = 1, 2, \ldots, n, \\
\mu_t &= \mu_{t-1} + \beta + \eta_t, \\
\psi_t &= \phi_1 \psi_{t-1} + \phi_2 \psi_{t-2} + \kappa_t, \\
\begin{pmatrix} \eta_t \\ \kappa_t \end{pmatrix} &\sim \text{NID}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_\eta & \sigma_{\eta\kappa} \\ \sigma_{\eta\kappa} & \sigma^2_\kappa \end{pmatrix} \right], & \sigma_{\eta\kappa} = r\sigma_\eta\sigma_\kappa.
\end{aligned}
\tag{11}
$$

It should be noticed that the trend and cycle disturbances are allowed to be contemporaneously correlated, with r being the correlation coefficient. The reduced form of model (11) is the ARIMA(2,1,2) model: $\Delta y_t = \beta + \frac{\theta(L)}{\phi(L)}\xi_t, \quad \xi_t \sim \text{NID}(0, \sigma^2)$, where $\theta(L) = 1 + \theta_1 L + \theta_2 L^2$ and $\phi(L) = 1 - \phi_1 L - \phi_2 L^2$. The structural form is exactly identified, both it and the reduced form have six parameters. The orthogonal trend cycle decomposition considered by Clark (1987) imposes the overidentifying restriction $r = 0$.

We estimate this model for the U.S. GDP series using the sample period 1947.1-2006.4. For comparison we also fit an unrestricted ARIMA(2,1,2) model and the restricted version imposing $r = 0$, which will be referred to henceforth as the Clark model. Estimation of the unknown parameters is carried out by frequency domain maximum likelihood estimation; see Nerlove, Grether and Carvalho (1995) and Harvey (1989, sec. 4.3) for the derivation of the likelihood function and the discussion on the nature of the approximation involved. Given the availability of the differenced observations $\Delta y_t, t = 1, 2, \ldots, n$, and denoting by $\omega_j = 2\pi j/n$, $j = 0, 1, \ldots, (n-1)$, the Fourier frequencies, the Whittle's likelihood is defined as follows:

$$
\text{loglik} = -\frac{n}{2}\ln 2\pi - \frac{1}{2}\sum_{j=0}^{n-1}\left[\log f(\omega_j) + \frac{I(\omega_j)}{f(\omega_j)}\right], \tag{12}
$$

where $I(\omega_j)$ is the sample spectrum,

$$
I(\omega_j) = \frac{1}{2\pi}\left[c_0 + 2\sum_{k=1}^{n-1} c_k \cos(\omega_j k)\right],
$$

c_k is the sample autocovariance of Δy_t at lag k, and $f(\omega_j)$ is the parametric spectrum of the implied stationary representation of the MNZ model, $\Delta y_t = \beta + \eta_t + \Delta\psi_t, t = 1, \ldots, n$, evaluated at the Fourier frequency ω_j. In particular,

$$
f(\omega) = f_{\Delta\mu}(\omega) + f_{\Delta\psi}(\omega) + f_{\Delta\mu,\Delta\psi}(\omega),
$$

with

$$
f_{\Delta\mu}(\omega) = \frac{\sigma^2_\eta}{2\pi}, \quad f_{\Delta\psi}(\omega) = \frac{1}{2\pi}\frac{2(1-\cos\omega)\sigma^2_\kappa}{\phi(e^{-\imath\omega})\phi(e^{\imath\omega})}, \quad f_{\Delta\mu,\Delta\psi}(\omega) = \frac{(1-e^{-\imath\omega})\phi(e^{\imath\omega}) + (1-e^{\imath\omega})\phi(e^{-\imath\omega})}{2\pi\phi(e^{-\imath\omega})\phi(e^{\imath\omega})} r\sigma_\eta\sigma_\kappa,
$$

$e^{-\imath\omega} = \cos\omega - \imath\sin\omega$, where $\imath$ is the imaginary unit, is the complex exponential, and $\phi(e^{-\imath\omega}) = 1 - \phi_1 e^{-\imath\omega} - \phi_2 e^{-2\imath\omega}$. The last term is the cross-spectrum of $(\Delta\psi_t, \Delta\mu_t)$, and of course it vanishes if $r = 0$. For the Clark model the parametric spectrum is given by the above expression with $f_{\Delta\mu,\Delta\psi}(\omega) = 0$, whereas for the unrestricted ARIMA(2,1,2) it is given by $f(\omega) = \sigma^2\theta(e^{-\imath\omega})\theta(e^{\imath\omega})[\phi(e^{-\imath\omega})\phi(e^{\imath\omega})]^{-1}$.

Figure 1 displays the quarterly growth rates, Δy_t, of U.S. GDP in the first panel. The next panel plots the profile likelihood for the correlation parameter against the value of r in [-1,1] and shows the presence of two modes, the first around -.9 and the second around zero. The parameter estimates, along with their estimated standard errors, and the value of the maximised likelihood, are reported in table 1[2]. It should be noticed that the unrestricted ARIMA(2,1,2) is exactly coincident with the reduced form of the MNZ model, as the two models yield the same likelihood and the AR and MA parameters are the mapping of the structural parameters. Secondly, the estimated correlation coefficient is high and negative (-0.93) and the likelihood ratio test of the hypothesis $r = 0$ has a p-value equal to 0.097. MNZ interpret the negative disturbance correlation as strengthening the case for the importance of real shocks in the macro economy: real shocks tend to shift the long run path of output, so short term fluctuations will largely reflect adjustments toward a shifting trend if real shocks play a dominant role.

Table 1: Frequency Domain Maximum Likelihood Estimation results for quarterly U.S. real GDP, 1947.1-2006.4

	ARIMA	MNZ	Clark
ϕ_1	1.34 (0.07)	1.34 (0.07)	1.49 (0.05)
ϕ_2	-0.76 (0.16)	-0.76 (0.16)	-0.56 (0.11)
θ_1	-1.08 (0.11)		
θ_2	0.59 (0.20)		
σ^2	0.8224 (0.08)		
r		-0.93 (0.28)	0(r)
σ^2_η		1.2626 (0.08)	0.3478 (0.15)
σ^2_κ		0.3556 (0.33)	0.4120 (0.16)
loglik	-315.76	-315.76	-317.14

The bottom left panel of figure 1 displays the sample spectrum $I(\omega_j)$ of Δy_t along with the estimated parametric spectral densities for the MNZ model (which is of course coincident with that of the ARIMA(2,1,2) model) and the Clark restricted model ($r = 0$). For the ARIMA(2,1,2) and the MNZ models the roots of the AR polynomial are a pair of complex conjugates that imply a spectral peak for Δy_t at the frequency 0.68, corresponding to a period of 9 quarters. As a matter of fact, a dominant feature of Δy_t is the presence of a cyclical component with a period of roughly two years. On the other hand, the spectral density implied by the Clark model peaks at the frequency 0.09, corresponding to a period of 68 quarters (i.e., a medium-run cycle).

A closer inspection of the sample spectrum reveals the presence of two consecutive periodogram ordinates, corresponding to a cycle of roughly two years, that are highly influential on the estimation results (they are circled in figure 1). It is indeed remarkable that when these are not used in the estimation, the correlation coefficient turns positive ($\hat{r} = 0.35$). The last panel of the figure presents the leave-two-out cross-validation estimates of the correlation coefficient, which are obtained by maximising Whittle's likelihood after deleting two consecutive periodogram ordinates at the frequencies ω_j and ω_{j+1}. This is a special case of weighted likelihood estimation, where each summand in (12) receives a weight equal to 1 if the frequency ω_j is retained and 0 if it is deleted.

The real time and the smoothed estimates of the cyclical component arising from the MNZ model, $\tilde{\psi}_{t|t} = \mathrm{E}(\psi_t|Y_t)$ and $\tilde{\psi}_{t|n} = \mathrm{E}(\psi_t|Y_n)$, respectively, are reported in figure 2, along with the 95% interval estimates; here Y_t denotes the information available up to and including time t. The bottom panels display the weights $\mathrm{w}_{\psi,j}$ of the signal extraction filters $\sum_j \mathrm{w}_{\psi,j} L^j y_t$ that yield the cycle estimates in the two cases.

[2]All the computations in this chapter have been performed using Ox version 4, see Doornik (2006).

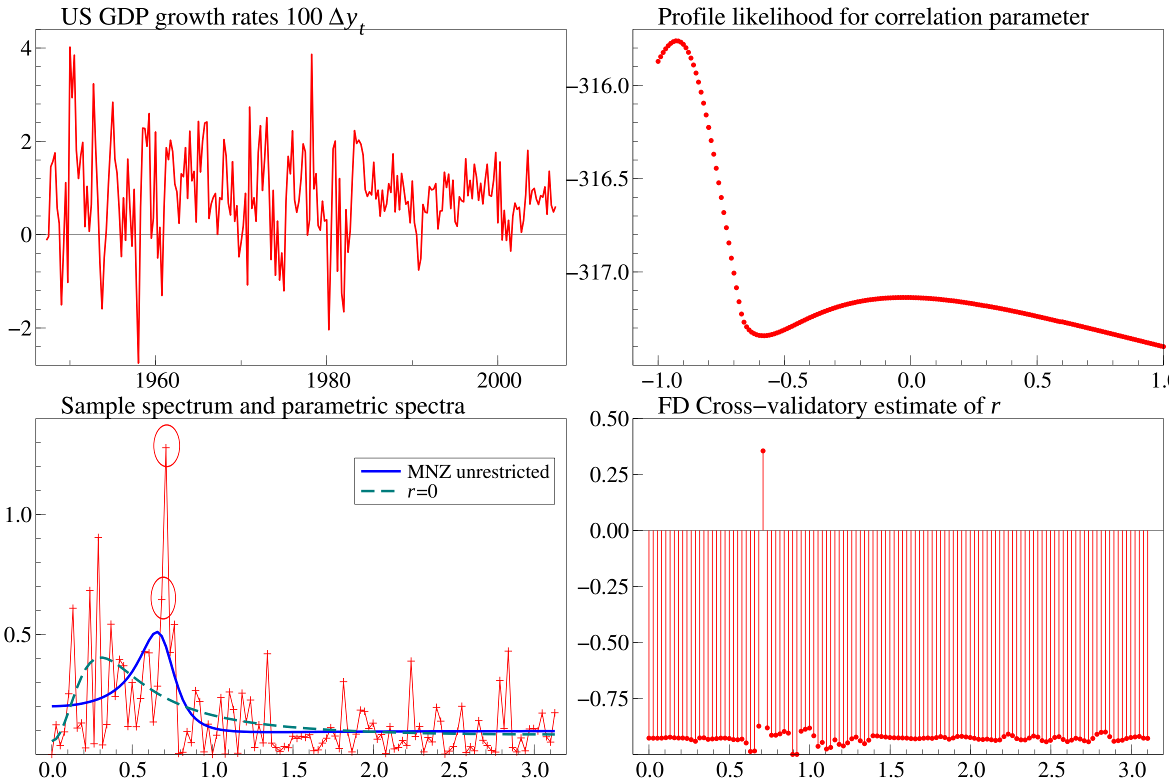

Figure 1: Quarterly U.S. real growth, 1947.2-2006.4. Sample spectrum and parametric spectral fit of trend-cycle model with correlated components.

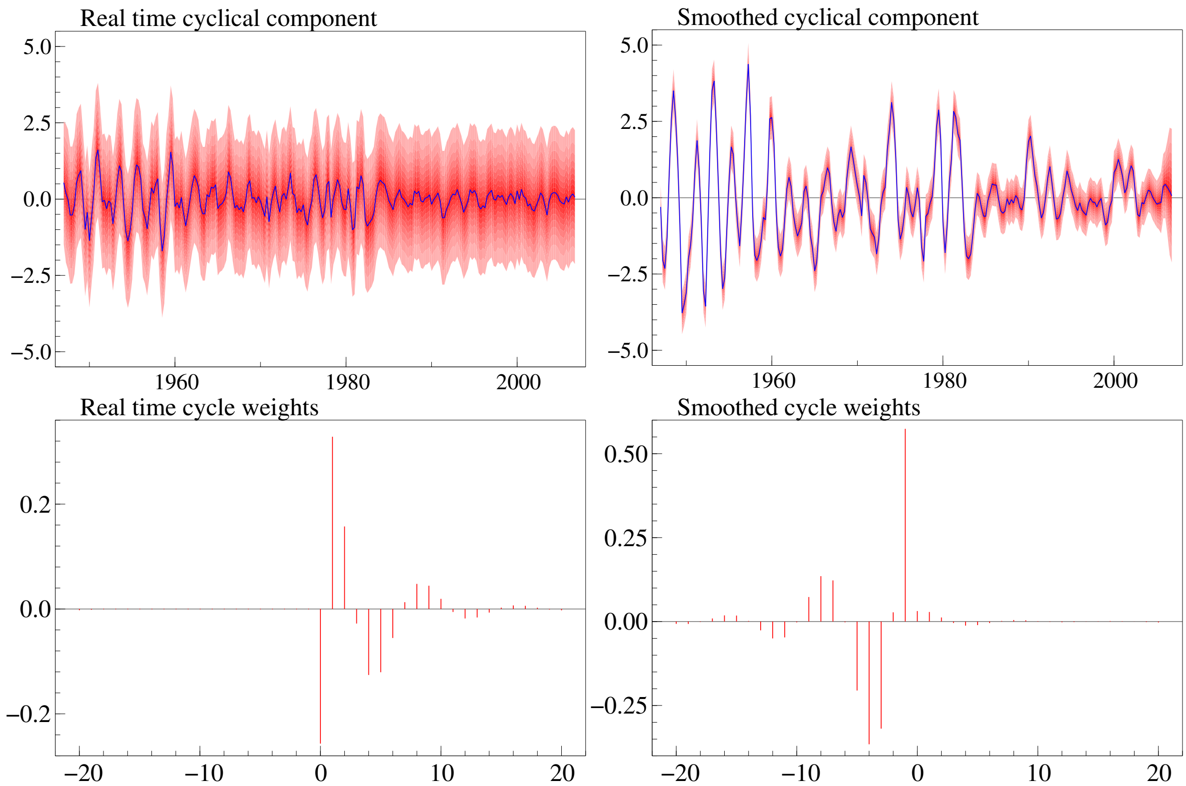

Figure 2: Trend-cycle decomposition with correlated disturbances. Real time and smoothed estimates of the cyclical components.

The real time estimates support the view that most of the variation in GDP is permanent, i.e., it should be ascribed to changes in the trend component, whereas little variance is attributed to the transitory component. In fact, the amplitude of $\tilde{\psi}_{t|t}$ is small and the interval estimates of ψ_t in real time are never significantly different from zero. When we analyse the smoothed estimates the picture changes quite radically: the cycle estimates are much more variable and there is a dramatic reduction in the estimation error variance, so that the contribution of the transitory component to the variation in GDP is no longer negligible. The real time estimates provide a gross underestimation of the cyclical component and are heavily revised as the future missing observations become available. As a matter of fact, the final estimates depend heavily on future observations, as can be seen from the pattern of the weights in the last panel of figure 2. That this behaviour is typical of the MNZ model when $\hat{r}$ is high and negative is documented in Proietti (2006a).

The real time estimates of the trend and cyclical components are coincident with the Beveridge and Nelson (1981, BN henceforth) components defined for the ARIMA(2,1,2) reduced form. The BN decomposition defines the trend component at time t as the value of the eventual forecast function at that time, or, equivalently, the value that the series would take if it were on its long run path (see also Brewer, 1979). For an ARIMA$(p, 1, q)$ process, this argument defines the trend as a random walk driven by the innovations $\xi_t = y_t - \mathrm{E}(y_t|Y_{t-1})$. Writing the ARIMA representation for y_t as $\Delta y_t = \beta + \psi(L)\xi_t$, $\psi(L) = \theta(L)/\phi(L)$, where $\phi(L)$ is a stationary AR polynomial of order p and $\theta(L)$ an invertible MA polynomial of order q, the BN decomposition can be written as: $y_t = m_t + c_t, t = 1, ..., n$, where m_t is the BN trend, and c_t is the cyclical component.

The trend is defined as $\lim_{l\to\infty}[\tilde{y}_{t+l|t} - l\beta]$, with $\tilde{y}_{t+l|t} = \mathrm{E}(y_{t+l}|Y_t)$. Writing $y_{t+l} = y_{t+l-1} + \beta + \psi(L)\xi_t$, taking the conditional expectation and rearranging, it is easily shown to give $m_t = m_{t-1} + \beta + \psi(1)\xi_t$, where $\psi(1) = \theta(1)/\phi(1)$ is the persistence parameter, as it measures the fraction of the innovation at time t that is retained in the trend. In terms of the observations, $m_t = \mathrm{w}_m(L)y_t$, where $\mathrm{w}_m(L)$ is the one-sided filter

$$w_m(L) = \frac{\psi(1)}{\psi(L)} = \frac{\theta(1)}{\phi(1)}\frac{\phi(L)}{\theta(L)}.$$

The sum of the weights is one, that is $w_m(1) = 1$.

The *transitory component* is defined residually as $c_t = y_t - m_t = \psi^*(L)\xi_t$, where $\Delta\psi^*(L) = \psi(L) - \psi(1)$. Alternative representations in terms of the observations y_t and of the innovations ξ_t are, respectively:

$$c_t = \frac{\phi(1)\theta(L) - \theta(1)\phi(L)}{\phi(1)\theta(L)}y_t, \qquad c_t = \frac{\phi(1)\theta(L) - \theta(1)\phi(L)}{\phi(1)\phi(L)\Delta}\xi_t. \tag{13}$$

The first expression shows that the weights for the extraction of the cycle sum to zero. Since $\phi(1)\theta(L) - \theta(1)\phi(L)$ must have a unit root, we can write $\phi(1)\theta(L) - \theta(1)\phi(L) = \Delta\vartheta(L)$, and substituting this into (13), the ARMA representation for this component can be established as $\phi(L)c_t = \vartheta(L)[\phi(1)]^{-1}\xi_t$. As the order of $\vartheta(L)$ is $\max(p, q) - 1$, the cyclical component has a stationary ARMA$(p, \max(p, q) - 1)$ representation. For the ARIMA(2,1,2) model fitted to U.S. GDP, the cycle has the ARMA(2,1) representation:

$$\phi(L)c_t = (1 + \vartheta L)\left[1 - \frac{\theta(1)}{\phi(1)}\right]\xi_t, \qquad \vartheta = -\frac{\phi_2\theta(1) + \theta_2\phi(1)}{\phi(1) - \theta(1)}. \tag{14}$$

It is apparent that the two components are driven by the innovations, ξ_t; the fraction $\theta(1)/\phi(1)$, known as *persistence*, is integrated in the trend, and its complement to 1 drives the cycle. The sign of the correlation between the trend and the cycle disturbances is provided by the sign of $\phi(1) - \theta(1)$; when persistence is less (greater) than one then trend and cycle disturbances are positively (negatively) and perfectly correlated.

2.6 Model–based band-pass filters

As we said before, macroeconomic time series such as GDP do not usually admit the decomposition $y_t = \mu_t + \psi_t$, with ψ_t being a purely irregular component; nevertheless, applications of the class of filters (6) is widespread, as the popularity of the Hodrick-Prescott filter testifies. However, when the available series y_t cannot be modelled as (2) it is not immediately clear how the components should be defined and how inferences about them should be made. In particular, the Kalman filter and the associated smoothing algorithms no longer provide the minimum mean square estimators of the components nor their mean square error. The discussion of model-based band-pass filtering in a more general setting will be the theme of this section.

The trend-cycle decompositions dealt with in the two previous sections are models of economic fluctuations, such that the components are driven by random disturbances which are propagated according to a transmission mechanism. In this section we start from a reduced form model (as in the case of the BN decomposition) and define parametric trend-cycle decompositions that are less loaded with structural interpretation, since they just represent the low-pass and the high-pass components in the series. The aim is to motivate and extend the use of signal extraction filters of the class (6) to a more general and realistic setting than (5). For this approach to the definition of band-pass filters see Gòmez (2001) and Kaiser and Maravall (2005). The following treatment is based on Proietti (2004).

Let y_t denote a univariate time series with ARIMA(p, d, q) representation, that we write

$$\phi(L)(\Delta^d y_t - \beta) = \theta(L)\xi_t, \quad \xi_t \sim \text{NID}(0, \sigma^2),$$

where c is a constant, $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$ is the AR polynomial with stationary roots, and $\theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$ is invertible. We are going to exploit the fundamental idea that we can uniquely decompose the WN disturbance ξ_t into two orthogonal stationary processes as follows:

$$\xi_t = \frac{(1+L)^r \zeta_t + (1-L)^m \kappa_t}{\varphi(L)}, \tag{15}$$

where ζ_t and κ_t are two mutually and serially independent Gaussian disturbances, $\zeta_t \sim \text{NID}(0, \sigma^2)$, $\kappa_t \sim \text{NID}(0, \lambda\sigma^2)$, and

$$|\varphi(L)|^2 = \varphi(L)\varphi(L^{-1}) = |1+L|^{2r} + \lambda|1-L|^{2m}. \tag{16}$$

We assume that λ is known. Equation (16) is the spectral factorisation of the lag polynomial on the right hand side; the existence of the polynomial $\varphi(L) = \varphi_0 + \varphi_1 L + \cdots + \varphi_{q^*} L^{q^*}$, of degree $q^* = \max(m, r)$, is guaranteed by the fact that the Fourier transform of the right hand side is never zero over the entire frequency range; see Sayed and Kailath (2001) for details.

According to (15), for given values of λ, m and r, the innovation ξ_t is decomposed into two ARMA(2,2) processes, characterised by the same AR polynomial, but by different MA components. The first component will drive the low–pass component of y_t and its spectral density is proportional to $\sigma^2 \mathrm{w}_\mu(\omega)$, where $\mathrm{w}_\mu(\omega)$ is the gain of the filter (6). If $r > 0$ the MA representation is non–invertible at the π frequency. Notice that, as m and r increase, the transition from the pass–band to the stop–band is sharper.

Substituting (15)-(16) into the ARIMA representation, the series can be decomposed into two orthogonal components:

$$\begin{array}{rcll} y_t &=& \mu_t + \psi_t, & \\ \phi(L)\varphi(L)(\Delta^d \mu_t - \beta) &=& (1+L)^r \theta(L)\zeta_t, & \zeta_t \sim \text{NID}(0, \sigma^2) \\ \phi(L)\varphi(L)\psi_t &=& \Delta^{m-d}\theta(L)\kappa_t, & \kappa_t \sim \text{NID}(0, \lambda\sigma^2). \end{array} \tag{17}$$

The trend or low-pass component has the same order of integration as the series (regardless of m), whereas the cycle or high-pass component is stationary provided that $m \geq d$, which will be assumed throughout.

Given the availability of a doubly infinite sample, the Wiener-Kolmogorov estimators of the components are $\tilde{\mu}_t = \mathrm{w}_\mu(L)y_t$ and $\tilde{\psi}_t = [1 - \mathrm{w}_\mu(L)]y_t$, where the impulse response function of the optimal filters is given by (6). Hence, the signal extraction filter for the central data points will continue to be represented by (6), regardless of the properties of y_t, but this is the only feature that is invariant to the nature of the time series and its ARIMA representation. The mean square error of the smoothed components, as a matter of fact, depends on the ARIMA model for y_t. In finite samples, the estimators and their mean square errors will be provided by the Kalman filter and smoother associated with the model (17), and thus will depend on the ARIMA model for y_t.

Band-pass filters can also be constructed from the principle of decomposing the low-pass component in (17). Let us consider fixed values of m and r and two cutoff frequencies, ω_{c1} and $\omega_{c2} > \omega_{c1}$, with corresponding values of the smoothness parameter λ_1 and λ_2, determined according to (7). Obviously $\lambda_1 > \lambda_2$. The trend-cycle decomposition corresponding to the triple m, r, λ_2 (or, equivalently, m, r, ω_{c2}), is as in (17):

$$\begin{array}{rcll} y_t & = & \mu_{2t} + \epsilon_t, & \\ \Delta^d \mu_{2t} & = & \beta + \frac{(1+L)^r}{\varphi_2(L)} \frac{\theta(L)}{\phi(L)} \zeta_{2t}, & \zeta_{2t} \sim \text{NID}(0, \sigma^2) \\ \epsilon_t & = & \frac{(1-L)^m}{\varphi_2(L)} \frac{\theta(L)}{\Delta^d \phi(L)} \kappa_{2t}, & \kappa_{2t} \sim \text{NID}(0, \lambda_2 \sigma^2) \end{array} \tag{18}$$

with $|\varphi_2(L)|^2 = |1 + L|^{2r} + \lambda_2 |1 - L|^{2m}$.

We can similarly define the trend-cycle decomposition corresponding to the triple m, r, λ_1 (or, equivalently, m, r, ω_{c1}), $y_t = \mu_{1t} + \psi_t$. As $\lambda_1 > \lambda_2$ this decomposition features a lower cutoff frequency, ω_{c1}, thereby yielding a smoother trend. The components μ_{1t} and ψ_t are defined as in (18), with $\varphi_1(L)$, $\zeta_{1t} \sim \text{NID}(0, \sigma^2)$ and $\kappa_{1t} \sim \text{NID}(0, \lambda_1 \sigma^2)$ replacing respectively $\varphi_2(L)$, ζ_{2t} and κ_{2t}. The polynomial $\varphi_1(L)$ is such that $|\varphi_1(L)|^2 = |1 + L|^{2r} + \lambda_1 |1 - L|^{2m}$.

The low-pass component, μ_{2t}, can, in turn, be decomposed using the orthogonal decomposition of the disturbance ζ_{2t}:

$$\zeta_{2t} = \frac{\varphi_2(L)}{\varphi_1(L)} \zeta_{1t} + \frac{(1-L)^m}{\varphi_1(L)} \kappa_{1t} \tag{19}$$

with

$$\zeta_{1t} \sim \text{NID}(0, \sigma^2), \quad \kappa_{1t} \sim \text{NID}\left(0, (\lambda_1 - \lambda_2)\sigma^2\right), \text{E}(\zeta_{1j}\kappa_{1t}) = 0, \forall j, t.$$

Under this setting, the spectrum of both sides of (19) is constant and equal to $\sigma^2 / 2\pi$.

Substituting (19) into (18), and writing $\mu_{2t} = \mu_{1t} + \psi_t$, enables y_t to be decomposed into three components, representing the low-pass (μ_{1t}), bandpass (ψ_t) and high-pass (ϵ_t) components, respectively.

$$\begin{array}{rcll} y_t & = & \mu_{1t} + \psi_t + \epsilon_t, & \\ \Delta^d \mu_{1t} & = & c + \frac{(1+L)^r}{\varphi_1(L)} \frac{\theta(L)}{\phi(L)} \zeta_{1t}, & \zeta_{1t} \sim \text{NID}(0, \sigma^2) \\ \psi_t & = & \frac{(1+L)^n (1-L)^m}{\varphi_1(L)\varphi_2(L)} \frac{\theta(L)}{\Delta^d \phi(L)} \kappa_{1t}, & \kappa_{1t} \sim \text{NID}\left(0, (\lambda_1 - \lambda_2)\sigma^2\right) \end{array} \tag{20}$$

and ϵ_t, given in (18), is the high-pass component of the decomposition (20). The model can be cast in state space form and the Kalman filter and smoother (see Appendix C) will provide the optimal estimates of the components and their standard errors.

Figure 3 shows the gain of an ideal band-pass filter and the Baxter and King filter. The dashed line is the gain of the model-based band-pass filter which is optimal for ψ_t in (20) using $m = r = 6$ and the two cut-off frequencies $\omega_{c1} = 2\pi/32$ (correponding to a period of 8 years for quarterly data) and $\omega_{c2} = 2\pi/6$ (1.5 years); such large values of the parameters yield a gain which is close to the ideal box-car function. The HP band-pass curve is the gain of the Wiener-Kolmogorov filter for extracting the component ψ_t in (20) with $m = 2, r = 0$, and ω_{c1}, ω_{c2} given above. In this case the leakage is larger but, as shown in Proietti (2004), taking large values of m and r is detrimental to the reliability of the end of sample estimates.

2.7 Applications of model-based filtering: band-pass cycles and the estimation of recession probabilities

We present two applications of the model-based filtering approach outlined in the previous section. Our first illustration deals with the estimation and the assessment of the reliability of the deviation cycle in U.S. GDP. The cycle is defined as the high-pass component extracting the fluctuations in the level of log GDP that have a periodicity smaller than ten years (40 quarters). To evaluate model uncertainty, we fit three models to the logarithm of GDP, namely a simple random walk, or ARIMA(0,1,0) model ($\hat{\sigma}^2 = 0.8570$), an ARIMA(1,1,0) model (the estimated

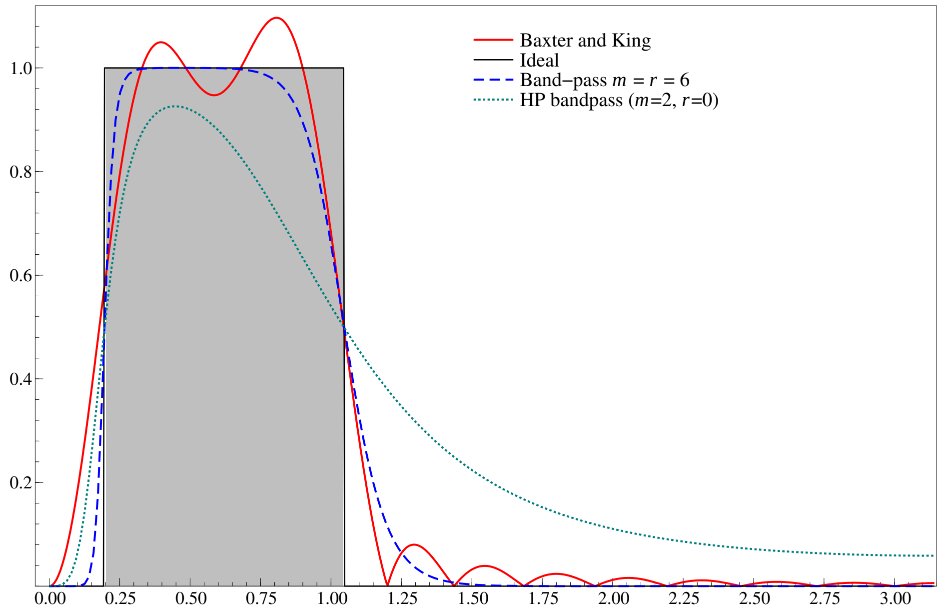

Figure 3: Gain function of the ideal business cycle band-pass filter, the Baxter and King filter and two model based filters.

first order autoregressive coefficient is $\hat{\phi} = 0.33$ and $\hat{\sigma}^2 = 0.8652$), and finally, we considered the ARIMA(2,1,2) model fitted in section 2.5, whose parameter estimates were reported in table 1.

The estimates of the low-pass component corresponding to the three models setting $m = 2, r = 0$ (and thus $\lambda = 1600$ and $\omega_c = 0.158279$) are displayed in the top right hand panel of figure 4, along with the Leser-HP cycle. The estimates for the three models are obtained as the conditional mean of ψ_t given the observations by applying the Kalman filter and smoother to the representation (17); the algorithm also provides their estimation error variance. It must be stressed that the Leser-HP filter is optimal for a restricted IMA(2,2) process and thus it does not yield the minimum mean square estimator of the cycle, nor its standard error. In general, also looking at the middle panel, which displays the estimated cycles for the last 12 years, the model-based estimates are almost indistinguishable, and are quite close to the Leser-HP cycle estimates in the middle of the sample. Large differences with the latter arise at the beginning, where the low-pass component had greater amplitude, and at the end of the sample period.

The particular model that is chosen matters little as far as the point estimates of ψ_t are concerned. Nevertheless, it is relevant for the assessment of the accuracy of the estimates, as can be argued from the right middle panel of the figure, which shows the estimation error variance, $\mathrm{Var}(\psi_t|Y_n)$ for the three models of US GDP. It is also evident that the standard errors obtained for the Leser-HP filter would underestimate the uncertainty of the estimates.

We conclude this first illustration by estimating the deviation cycle as a band-pass component, assuming that the true model is the ARIMA(2,1,2) and using the cut-off frequencies $\omega_{c1} = 2\pi/32$, $\omega_{c2} = 2\pi/6$, and the values $m = 2, r = 0$; as a consequence, the component ψ_t in (20) selects all the fluctuations in a range of periodicity that goes from one and a half years (6 quarters) to 8 years (32 quarters). The gain of the filter is displayed in figure 3. The estimates of ψ_t are compared to the Baxter and King cycle in the bottom left panel of figure 4 and to the corresponding high-pass estimates $(\psi_t + \epsilon_t)$. With respect to the BK cycle, the estimates are

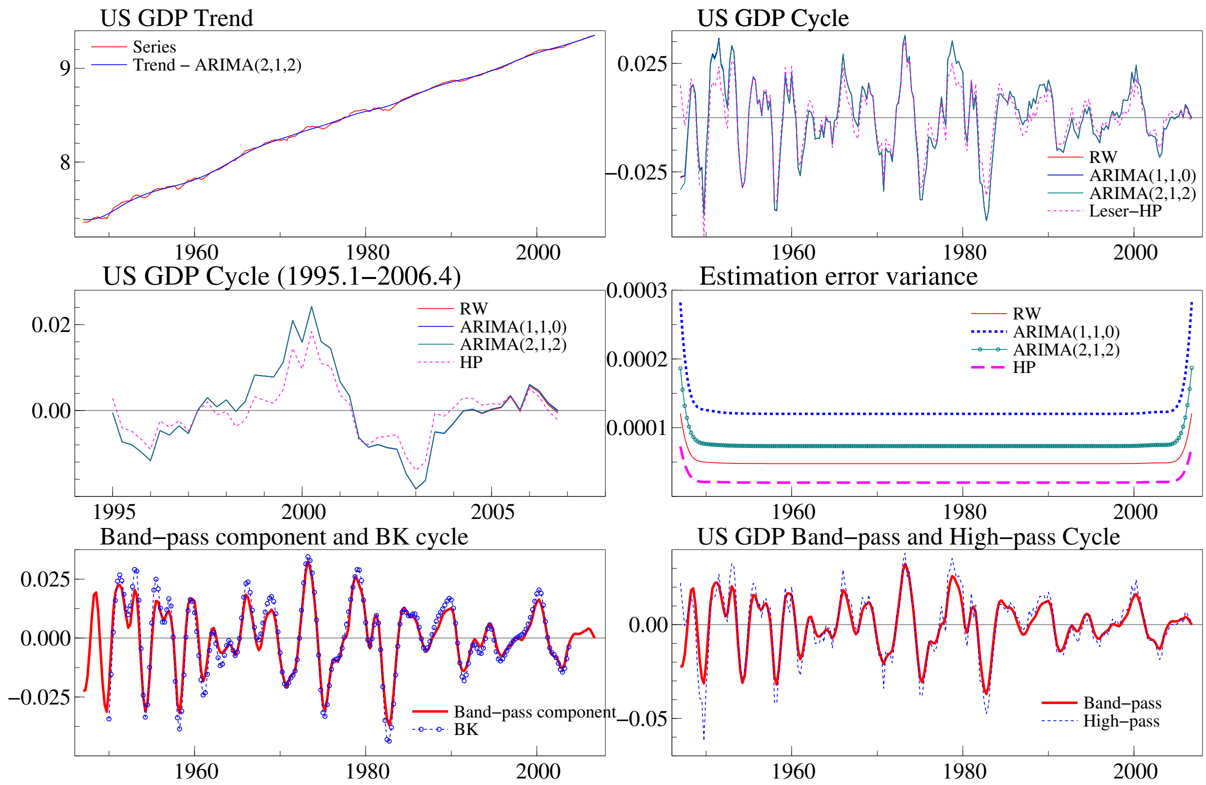

Figure 4: Model-based filtering. Estimates of the low-pass component (using the ARIMA(2,1,2) model) and of the high-pass and band-pass components in U.S. GDP, and their comparison with the Leser-HP cycle and the Baxter and King cycle.

available also in real time.

The conclusion is that model-based filtering improves the quality of the estimated low-pass component, providing estimates at the boundary of the sample period that are automatically adapted to the series under investigation, and enables the investigator to assess the reliability of the estimates (conditional on a particular reduced form).

The second application deals with assessing the uncertainty in estimating the business cycle chronology. According to the classical definition, the business cycle is a recurrent sequence of expansions and contractions in the aggregate level of economic activity; see Burns and Mitchell (1946, p. 3). Dating the business cycle amounts to establishing a set of reference dates that mark the phases or states of the economy. Usually two phases, recessions and expansions, are considered, that are delimited by peaks and troughs in economic activity. Dating is carried out by an algorithm, such as that due to Bry and Boschan (1971), or that proposed by Artis et al. (2004), which aims at estimating the location of turning points, enforcing the alternation of peaks and troughs and minimum duration ties for the phases and the full cycle. Downturns and upturns have to be persistent to qualify as cycle phases; thus, they need to fulfill minimum duration constraints, such as at least two quarters for each phase; moreover, to separate it from seasonality, a complete sequence, recession-expansion or expansion-recession, i.e. a full cycle, has to last longer than one year. Depth restrictions, motivated by the fact that only major fluctuations qualify for the phases, should also be enforced.

An integral part of the dating algorithm is prefiltering, which is necessary in order to isolate the fluctuations in the series with period greater than the minimum cycle duration. For instance, in the quarterly case we need to abstract from all the fluctuations with periodicity less than 5 quarters, i.e., from high frequency fluctuations that do not satisfy the minimum cycle duration. En lieu of the ad hoc and old fashioned moving averages adopted by Bry and Boschan, one can use model based low-pass signal extraction filters.

The advantages are twofold: on the one hand it is possible to select the cut-off frequency so as to match the minimum cycle duration; for instance, in our case $\omega_c = 2\pi/5$. Secondly, the uncertainty in dating arising from prefiltering can be assessed by Monte Carlo simulation, by means of an algorithm known as the simulation smoother, see de Jong and Shephard (1995), Durbin and Koopman (2002) and appendix C.4. This repeatedly draws simulated samples from the posterior distribution of the low-pass component with a cut-off frequency corresponding to 5 quarters, $\tilde{\mu}_t^{(i)} \sim \mu_t | Y_n$, so that by repeating the draws a sufficient number of times we can get Monte Carlo estimates of different aspects of the marginal and joint distribution of the low-pass component, intended here as the level of output devoid of all fluctuations with a periodicity smaller that 5 quarters.

Figure 5 plots the recession frequencies, i.e., the relative number of times each quarter was classified as a recessionary period. For this purpose 5000 draws from the conditional distribution of $\boldsymbol{\mu}|\mathbf{y}$ were extracted; each quarter was classified as recession or expansion according to the Artis et al. (2004) Markov chain dating algorithm. There is a close agreement with the NBER chronology, which is not based on GDP alone, and the last recession, which started in March 2001 and ended in October 2001, was really mild in terms of GDP; in fact, the recession frequency is only in one quarter greater than 0.5.

2.8 Ad-hoc filtering and the Slutsky-Yule effect

A filter is *ad hoc* when it is invariant to the properties of the time series under investigation. An instance is provided by the Leser-HP filter with a fixed smoothing parameter, and another example is the BK filter. The potential danger associated with an *ad hoc* cycle extraction filter is that the filtered series displays cyclical features that are absent from the original series. The risk of extracting spurious cycles is known in the time series literature as the Slutsky-Yule effect.

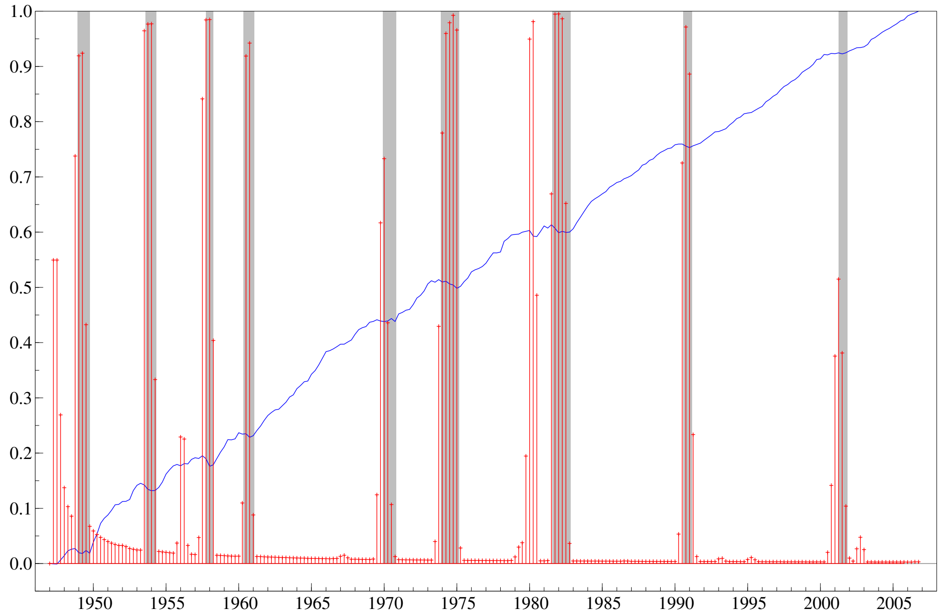

Figure 5: Relative number of times each quarter is classified as a recessionary period, using 5000 simulated samples. The shaded areas represent NBER recessions.

The distortionary effects of the Leser-HP filter have been discussed by King and Rebelo (1993), Harvey and Jäger (1993), and Cogley and Nason (1995). These authors document that, when the series to which the filter is applied is difference stationary (e.g., a random walk, or an integrated random walk), the detrended series can display spurious cyclical behaviour. As a matter of fact, the transfer function will display a distinctive peak at business cycle frequencies, which is only due to the leakage from the nonstationary component. Moreover, the filter seriously distorts the evidence for the comovements among detrended series.

The issue of spuriousness is problematic, at least, if not tautological. The main difficulty stems from the fact that it ties in with a more fundamental question concerning what is indeed the cycle in economic time series. If we adhere to the band–pass paradigm of viewing the cycle as consisting of those fluctuations within a give range of periodicity, than the case for spuriousness is much less compelling.

Another source of concern among practitioners, especially for the conduct of monetary policy, relates to the end of sample behaviour of the Leser-HP filter: the real time estimates would be subject to "end-of-sample bias", since they result from the application of a one-sided filter and will suffer from both phase shifts and amplitude distorsions. One has to separate two issues: as we hinted before, the IMA(2,2), for which the Leser-HP filter is the optimal filter, is usually misspecified for macroeconomic time series. As a result, the cycle estimates have no optimality properties. Model–based bandpass filtering is aimed at overcoming this limitation. Having said that, it is a fact of life that, for a correctly specified model, the optimal real time signal extraction filter will be one-sided and thus will produce phase-shifts and amplitude distortions. We will return to this issue in section 4.2.

3 Multivariate Models

Information on the output gap is contained in macroeconomic variables other than aggregate output, either because those variables provide alternative measures of production, or because they are functionally related to the output gap. In this section we start from the consideration of a bivariate model that, along with an output decomposition, includes an inflation equation. We then extend the model to include other variables, such as the unemployment rate and industrial production, and consider the estimation of a monthly model using quarterly observations on real GDP.

3.1 Bivariate models of real output and inflation

Price inflation carries relevant information for the output gap. The definition of the latter as an indicator of inflationary pressure and, correspondingly, of potential output as the level of output consistent with stable inflation, makes clear that a rigorous measurement can be made at least within a bivariate model of output and inflation, embodying a Phillips curve relationship. The Phillips curve establishes a relation between the nominal price or wage inflation rate, Δp_t, where, for instance, p_t is the logarithm of the consumer price index (CPI), and an indicator of excess demand, typically the output gap (ψ_t).

A general specification is the following:

$$\delta(L)\Delta p_t = c + \theta_\psi(L)\psi_t + \boldsymbol{\gamma}(L)'\mathbf{x}_t + \xi_{pt}, \tag{21}$$

where c is a constant, x_t denotes a set of exogenous supply shocks, such as changes in energy prices and terms of trade, and ξ_{pt} is WN. Often the restriction is imposed that the sum of the AR coefficients on lagged inflation is unity, $\delta(L) = \Delta\delta^*(L)$, where $\delta^*(L)$ is a stationary AR polynomial; the gap enters the equation with more than one lag to capture also the role of the change in demand, since we can rewrite $\theta_\psi(L) = \theta_\psi(1) + \Delta\theta^*_\psi(L)$. This is known as Gordon's "triangle" model of inflation, see Gordon (1997), since it features the three main driving forces: inertia (or inflation persistence, via $\delta(L)$), endogenous demand shocks (via ψ_t), and exogenous supply shocks (via $\mathbf{x}_t$). If $\delta(L)$ has a unit root and $\theta_\psi(1) \neq 0$ the output gap has permanent effects on the inflation rate. If, instead, $\theta_\psi(1) = 0$, then the output gap is neutral in the long run and the inflation rate shares a common cycle in the levels with output. Harvey *et al.* (2007) consider the Bayesian estimation of a bivariate model of output and inflation, where the cycle in inflation is driven by the output gap plus an idiosyncratic cycle.

Kuttner (1994) estimated potential output and the output gap for the U.S. using a bivariate model of real GDP and CPI inflation. The output equation was specified as in the Clark (1987) model, i.e., $y_t = \mu_t + \psi_t$, such that potential output is a random walk with drift and the output gap is an AR(2) process driven by orthogonal disturbances. The equation for the inflation rate is a variant of Gordon's triangle model:

$$\Delta p_t = c + \gamma\Delta y_{t-1} + \theta_\psi\psi_{t-1} + v(L)\xi_{pt},$$

according to which the inflation rate is linearly related to the lagged output gap and to lagged GDP growth; inflation persistence is captured by the MA feature, $v(L)\xi_{pt}$, where the disturbance ξ_{pt} is allowed to be correlated with the output gap disturbance, κ_t. The inclusion of lagged real growth is not formally justified by Kuttner, and the correlation between ξ_{pt} and κ_t makes the dynamic relationship between the output gap and inflation more elaborate than it appears at first sight (for instance, inflation depends on the contemporaneous value of the gap). Moreover, permanent shocks are allowed to drive inflation via the term $\Delta y_{t-1} = \beta + \eta_{t-1} + \Delta\psi_{t-1}$, so that it cannot be maintained that μ_t is the noninflationary level of output. Planas et al. (2007) consider the Bayesian estimation of Kuttner's bivariate model, with the only variant being that the MA feature is replaced by an autoregressive feature: $\delta(L)\Delta p_t = c + \gamma\Delta y_{t-1} + \theta_\psi\psi_{t-1} + \xi_{pt}$.

Gerlach and Smets (1999) again use a bivariate model of output and inflation, but the output gap generating equation takes the form of an aggregate demand equation featuring the lagged real interest rate as an explanatory variable. The inflation equation is specified as in (21) with $\delta(L) = \Delta$.

The Gordon triangle model may be interpreted as a reduced form of a structural model of inflation that embodies expectations; the presence of lagged inflation in the specification reflects backward looking inflation expectations. In the New Keynesian approach the Phillips curve is forward looking, as inflation depends on expected future inflation. Domenech and Gómez (2006) estimate a multivariate model of output fluctuations including a forward looking Phillips curve specified as follows:

$$\Delta p_t = c + \delta \mathrm{E}(\Delta p_{t+1}|\mathcal{F}_t) + \theta_\psi(L)\psi_t + \xi_{pt},$$

where $\mathcal{F}_t$ is the information set at time t. Basistha and Nelson (2007) estimate a bivariate model of output and inflation where the output equation features the MNZ decomposition with correlated components and in the inflation equation survey based expectations replace $\mathrm{E}(\Delta p_{t+1}|\mathcal{F}_t)$.

3.2 A bivariate quarterly model of output and inflation for the U.S.

This section is devoted to the estimation of a bivariate model for U.S. quarterly real GDP and the quarterly rate of inflation Δp_t, where p_t is the logarithm of quarterly CPI for the U.S, using data from the first quarter of 1950 to the fourth quarter of 2006. The KPSS test conducted on the inflation series leads to the rejection of the null of stationarity against a random walk for all the values of the lag truncation parameter up to 5; if a linear trend is considered and stationarity is tested against a random walk with drift, then the null is rejected also for much higher values of the lag truncation parameter. In the sequel, inflation will be taken to be integrated of order one. The model has the following specification:

$$\begin{array}{rcll} y_t & = & \mu_t + \psi_t, & t = 1, \dots, n, \\ \mu_t & = & \mu_{t-1} + \beta_t + \eta_t, & \eta_t \sim \mathrm{NID}(0, \sigma^2_\eta) \\ \psi_t & = & \phi_1\psi_{t-1} + \phi_2\psi_{t-2} + \kappa_t, & \kappa_t \sim \mathrm{NID}(0, \sigma^2_\kappa) \\ & & & \\ \Delta p_t & = & \tau_t + \varepsilon_{pt} & \varepsilon_{pt} \sim \mathrm{NID}(0, \sigma^2_{p\varepsilon}) \\ \tau_t & = & \tau_{t-1} + \theta_\psi(L)\psi_t + \eta_{\tau t} & \eta_{\tau t} \sim \mathrm{NID}(0, \sigma^2_{\tau\eta}); \end{array} \qquad (22)$$

where η_t, κ_t, ε_{pt},and κ^*_t are mutually independent.

The output equation is the usual decomposition into orthogonal components; the inflation equation is a decomposition into a core component, τ_t, and a transitory one. The changes in the core component are driven by the output gap and by the idiosyncratic disturbances $\eta_{\tau t}$. The lag polynomial $\theta_\psi(L) = \theta_{\psi 0} + \theta_{\psi 1}L$ can be rewritten as $\theta_\psi(1) - \theta_{\psi 1}\Delta$, which enables us to isolate the level effect of the gap from the change effect, which we expect to be positive, that is we expect $\theta_{\psi 1} < 0$. If $\theta_\psi(1) = 0$, the inflation equation can be rewritten $\Delta p_t = \tau^*_t - \theta_{\psi 1}\psi_t + \varepsilon_t$, with $\Delta\tau^*_t = \eta_{\tau t}$, so that output and inflation would share a common cycle.

We also extend the specification of model (22) to take into account an important stylised fact, known as the "great moderation" of the business cycle, and which consists of a substantive reduction in the volatility of GDP growth. This feature is visible from the plot of Δy_t in figure 1. The date when the structural break in volatility occurred is identified as the first quarter of 1984 (see Kim and Nelson, 1999, McConnell and Perez-Quiros, 2000, and Stock and Watson, 2003).

Let S_t denote an indicator variable which takes the value 1 in the high volatility state (which we label regime a) and 0 in the low volatility state (regime b). The trend and cycle disturbance

Table 2: Maximum likelihood estimation results for bivariate models of quarterly U.S. log GDP (y_t) and the consumer price inflation rate (Δp_t), 1950.1-2006.4.

	Bivariate		Great Moderation	
	Parameter	Std. Error	Parameter	Std. Error
	y_t equation			
σ^2_η	0.33	0.14		
$\sigma^2_{\eta a}$			0.58	0.27
$\sigma^2_{\eta b}$			0.13	0.05
σ^2_κ	0.38	0.15		
$\sigma^2_{\kappa a}$			0.47	0.24
$\sigma^2_{\kappa b}$			0.06	0.04
ϕ_1	1.47	0.06	1.55	0.06
ϕ_2	-0.54	0.10	-0.60	0.09
	Δp_t equation			
$\sigma^2_{p\varepsilon}$	0.11	0.03	0.12	0.03
$\sigma^2_{\tau\eta}$	0.05	0.02	0.05	0.02
$\theta_{\psi 0}$	0.12	0.05	0.12	0.06
$\theta_{\psi 1}$	-0.10	0.05	-0.10	0.06
	Wald tests of restriction $\theta_\psi(1) = 0$			
	2.00		1.68	
loglik	-447.79		-415.53	

variances are time varying and the model will be specified as in (22) with

$$\eta_t \sim \mathrm{N}\left(0, S_t\sigma^2_{\eta a} + (1 - S_t)\sigma^2_{\eta b}\right), \quad \kappa_t \sim \mathrm{N}\left(0, S_t\sigma^2_{\kappa a} + (1 - S_t)\sigma^2_{\kappa b}\right).$$

This will be referred to as the GM specification. We shall consider two cases: (i) the sequence S_t is deterministic, taking the value 1 before 1984:1, and 0 thereafter; (ii) S_t is a random process, which we model as a first order Markov Chain with initial probability $p(S_0 = 1) = 1$, i.e., we know for certain that the process started in a high variance state, and transition probabilities $P(S_t = j|S_{t-1} = i) = \mathcal{T}_{ij}, i = 0, 1$, with $\mathcal{T}_{ij} = 1 - \mathcal{T}_{ii}$ for $j \neq i$.

3.2.1 Maximum likelihood estimation

The bivariate model and its GM extension under assumption (i) were estimated by maximum likelihood in the time domain. The likelihood is evaluated by the Kalman filter, see Appendix C for details. The parameter estimates and the associated standard errors are reported in table 2. The estimated trend and cycle disturbance variances are smaller after 1984:1 (regime b), as expected, and the likelihood ratio test of the homogeneity hypothesis, $H_0 : \sigma^2_{\eta a} = \sigma^2_{\eta b}, \sigma^2_{\kappa a} = \sigma^2_{\kappa b}$, clearly leads to a rejection. The roots of the AR polynomial for the output gap are complex and the loadings of core inflation on the output gap are significantly different from zero at the 5% level. The table also reports the Wald test for the null of long run neutrality of the output gap, $H_0 : \theta_\psi(1) = 0$, which is accepted under both specifications, with p-values equal to 0.16 and 0.19. The evidence is thus that the output gap has only transitory effects on the level of inflation.

Figure 6 displays the point and 95% interval estimates of the output gap and the core component of inflation for both specifications. It is interesting that the explicit consideration of the great moderation of volatility makes the estimates of the output gap after the 1984 break more precise. In interpreting this result, we must stress that the interval estimates make no

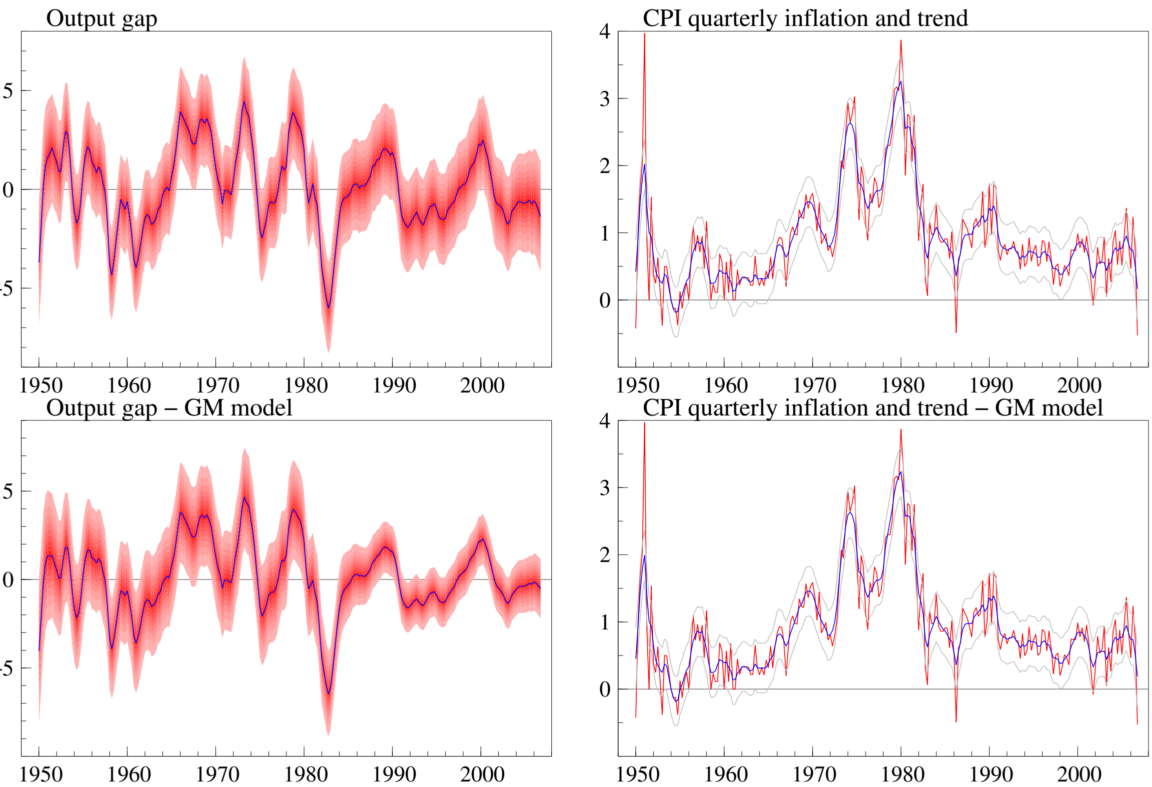

Figure 6: Estimates of the output gap and core inflation using the ML estimates of the parameters of the bivariate models of output and inflation under two specifications.

allowance for parameter uncertainty and for the uncertainty in dating the transition from the high volatility state to the low volatility one.

3.2.2 Bayesian estimation

Let us focus on the standard bivariate model (22) first and denote by $\mathbf{y}$ the stack of the observations $(y_t, \Delta p_t)$ for $t = 1, \ldots, n$, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_0', \ldots, \boldsymbol{\alpha}_n')'$, where the state vector at time t is $\boldsymbol{\alpha}_t = (\mu_t, \beta_t, \psi_t, \psi_{t-1}, \tau_t)$. Also, let $\boldsymbol{\mu}$, $\boldsymbol{\psi}$, $\boldsymbol{\eta}$, $\boldsymbol{\kappa}$, denote, respectively, the stack of potential output, the output gap, the disturbances η_t, and the cycle disturbances, where, for instance, $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_n)$, and let $\boldsymbol{\Xi} = [\phi_1, \phi_2, \sigma^2_\eta, \sigma^2_\kappa, \sigma^2_{p\varepsilon}, \sigma^2_{\tau\eta}, \theta_{\psi 0}, \theta_{\psi 1}]$ denote the vector of hyperparameters[3]. Notice that knowledge of $\boldsymbol{\alpha}$ implies knowledge of both the individual state components and the disturbances. Our main interest lies in aspects of the posterior marginal densities of the states given the observations, e.g., $f(\boldsymbol{\psi}|\mathbf{y})$ and $f(\boldsymbol{\Xi}|\mathbf{y})$: for instance $\mathrm{E}[h(\boldsymbol{\psi})] = \int h(\boldsymbol{\psi}) f(\boldsymbol{\psi}|\mathbf{y}) d\boldsymbol{\psi}$, for some function $h(\cdot)$ such as $h(\boldsymbol{\psi}) = \psi_t$. The computation of the integral is carried out by stochastic simulation: given a sample $\psi_t^{(i)}, i = 1, \ldots, M$, drawn from the posterior $f(\boldsymbol{\psi}|\mathbf{y})$, $\mathrm{E}[h(\boldsymbol{\psi})]$ is approximated by $M^{-1} \sum_i h\left(\psi_t^{(i)}\right)$. The required sample is obtained by Monte Carlo Markov Chain methods and, in particular, by a Gibbs sampling (GS) scheme that we now

[3]The slope parameter is included in the state vector; the transition equation is $\beta_t = \beta_{t-1}$, with β_0 being a diffuse parameter (see appendix C).

discuss in detail. This scheme produces correlated random draws from the joint posterior density $f(\boldsymbol{\alpha}, \boldsymbol{\Xi}|\mathbf{y})$, and thus from $f(\boldsymbol{\psi}|\mathbf{y})$, by repeatedly sampling an ergodic Markov chain whose invariant distribution is the target density; see Chib (2001) and the references therein.

This is achieved by the following iterative scheme. Specify an initial value $\boldsymbol{\alpha}^{(1)}, \boldsymbol{\Xi}^{(1)}$. For $i = 1, 2, \ldots, M$:

1. generate $\boldsymbol{\alpha}^{(i)} \sim f(\boldsymbol{\alpha}|\boldsymbol{\Xi}^{(i-1)}, \mathbf{y})$ using the simulation smoother, see Appendix C.4;
2. generate $\boldsymbol{\Xi}^{(i)} \sim f(\boldsymbol{\Xi}^{(i)}|\boldsymbol{\alpha}^{(i)}, \mathbf{y})$ This block is divided into smaller components, whose full conditional distribution is available for sampling. In particular,
   (a) Generate $(\phi_1^{(i)}, \phi_2^{(i)})'$ from the full conditional $(\phi_1, \phi_2)'|\boldsymbol{\psi}, \sigma_\kappa^{2(i-1)}$ (this distribution is conditionally independent of $\mathbf{y}$, given $\boldsymbol{\psi}$). Assuming a Gaussian prior distribution, $\mathrm{N}(\mathbf{m}_{\phi 0}, \boldsymbol{\Sigma}_{\phi 0})$, $(\phi_1, \phi_2)'|\boldsymbol{\psi}, \sigma_\kappa^{2(i-1)} \sim \mathrm{N}(\mathbf{m}_{\phi 1}, \boldsymbol{\Sigma}_{\phi 1})$ where, denoting $\boldsymbol{\chi}_{t-1} = (\psi_{t-1}^{(i-1)}, \psi_{t-2}^{(i-1)})'$,
   $$\boldsymbol{\Sigma}_{\phi 1} = \left(\boldsymbol{\Sigma}_{\phi 0}^{-1} + \frac{1}{\sigma_\kappa^{2(i-1)}} \sum_t \boldsymbol{\chi}_{t-1}\boldsymbol{\chi}_{t-1}'\right)^{-1}, \quad \mathbf{m}_{\phi 1} = \boldsymbol{\Sigma}_{\phi 1}\left(\boldsymbol{\Sigma}_{\phi 0}^{-1}\mathbf{m}_{\phi 0} + \frac{1}{\sigma_\kappa^{2(i-1)}} \sum_t \boldsymbol{\chi}_{t-1}\psi_t\right).$$
   The generations are repeated until a draw falls inside the stationarity region.
   (b) Generate $\sigma_\eta^{2(i)}$ from the full conditional inverse gamma (IG) distribution
   $$\sigma_\eta^2|\boldsymbol{\eta}^{(i-1)} \sim IG\left(\frac{v_\eta + n}{2}, \frac{\delta_\eta + \sum_t \eta_t^{(i-1)^2}}{2}\right)$$
   This assumes that the prior distribution is $\sigma_\eta^2 \sim IG(v_\eta/2, \delta_\eta/2)$.
   (c) Generate $\sigma_\kappa^{2(i)}$ from the full conditional IG distribution
   $$\sigma_\kappa^2|\boldsymbol{\kappa}^{(i-1)} \sim IG\left(\frac{v_\kappa + n}{2}, \frac{\delta_\kappa + \sum_t \kappa_t^{(i-1)^2}}{2}\right).$$
   This assumes that the prior distribution is $\sigma_\kappa^2 \sim IG(v_\kappa/2, \delta_k/2)$.
   (d) Generate $(\theta_{\psi 0}^{(i)}, \theta_{\psi 1}^{(i)})'$. Assuming the Gaussian prior $(\theta_{\psi 0}, \theta_{\psi 1})' \sim \mathrm{N}(\mathbf{m}_{\theta 0}, \boldsymbol{\Sigma}_{\theta 0})$, the full posterior is $(\theta_{\psi 0}, \theta_{\psi 1})'|\boldsymbol{\tau}, \sigma_{\tau\eta}^{2(i-1)} \sim \mathrm{N}(\mathbf{m}_{\theta 1}, \boldsymbol{\Sigma}_{\theta 1})$, where $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_n)$, and
   $$\boldsymbol{\Sigma}_{\theta 1} = \left(\boldsymbol{\Sigma}_{\theta 0}^{-1} + \frac{1}{\sigma_{\tau\eta}^{2(i-1)}} \sum_t \boldsymbol{\chi}_t\boldsymbol{\chi}_t'\right)^{-1}, \quad \mathbf{m}_{\phi 1} = \boldsymbol{\Sigma}_{\phi 1}\left(\boldsymbol{\Sigma}_{\theta 0}^{-1}\mathbf{m}_{\phi 0} + \frac{1}{\sigma_{\tau\eta}^{2(i-1)}} \sum_t \boldsymbol{\chi}_t\Delta\tau_t\right).$$
   (e) Generate $\sigma_{p\varepsilon}^{2(i)}$ from the full conditional IG distribution:
   $$\sigma_{p\varepsilon}^2|\boldsymbol{\varepsilon}_p^{(i-1)} \sim IG\left(\frac{v_\varepsilon + n}{2}, \frac{\delta_\varepsilon + \sum_t (\varepsilon_t^{(i-1)})^2}{2}\right).$$
   Here $\boldsymbol{\varepsilon}_p$ is the stack of the inflation equation measurement disturbances, and we assume the prior $\sigma_{p\varepsilon}^2 \sim IG(v_\varepsilon/2, \delta_\varepsilon/2)$.
   (f) Generate $\sigma_{\tau\eta}^{2(i)}$ from the full conditional IG distribution
   $$\sigma_{\tau\eta}^2|\boldsymbol{\eta}_\tau^{(i-1)} \sim IG\left(\frac{v_\tau + n}{2}, \frac{\delta_\tau + \sum_t \eta_{\tau t}^{(i-1)^2}}{2}\right),$$
   where $\boldsymbol{\eta}_\tau$ is the stack of the inflation equation core level disturbances, and we assume the prior $\sigma_{\eta\tau}^2 \sim IG(v_\tau/2, \delta_\tau/2)$.

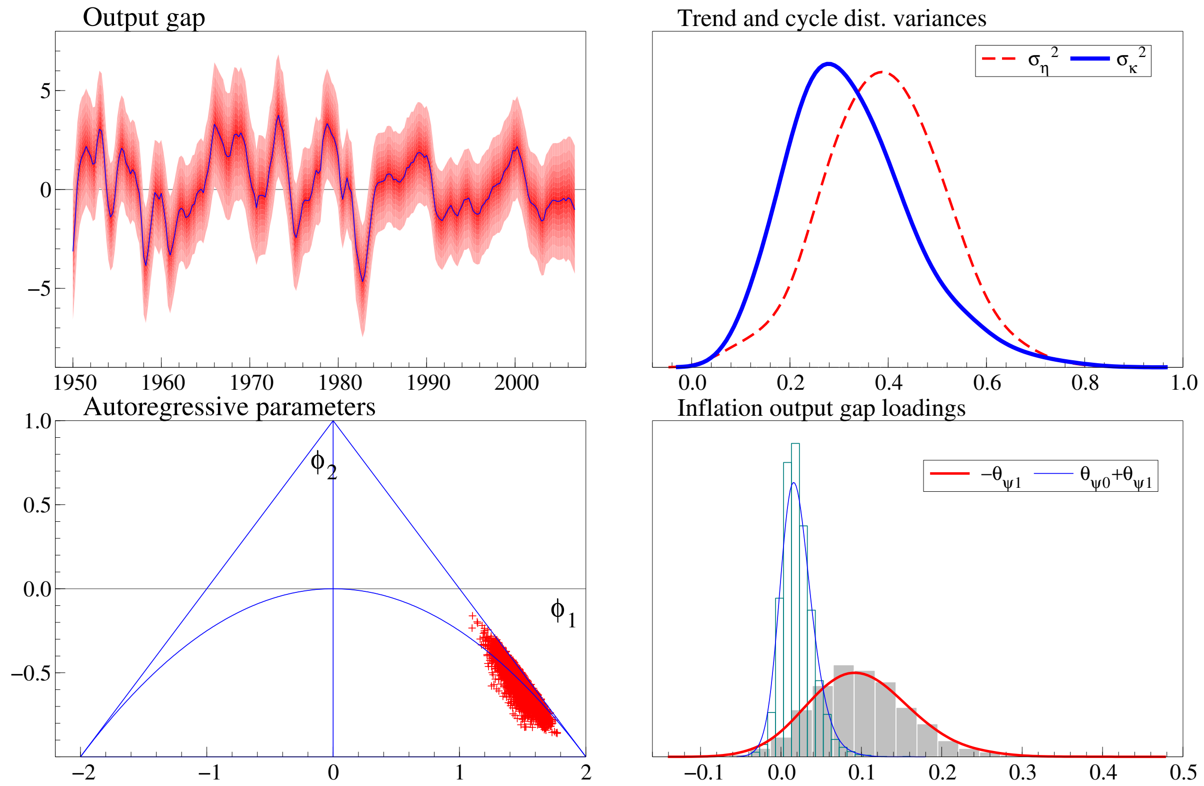

Figure 7: Bayesian estimation of the standard bivariate output gap model. Point and 95% interval estimates of the output gap; posterior densities of variance and loadings parameters; draws from the posterior of the AR parameters.

The above GS scheme defines a homogeneous Markov Chain such that the transition kernel is formed by the full conditional distributions and the invariant distribution is the unavailable target density.

The IG prior for the variance parameter is centred around the maximum likelihood estimate and is not very informative ($v_\eta = v_\kappa = v_\varepsilon = v_\tau = 4$, and $n = 426$); for the AR parameters and the loadings impose a standard normal prior. The number of samples is $M = 2000$ after a burn-in sample of size 1000. Figure 7 displays the posterior means and the 95% interval estimates of the output gap (first panel), along with a nonparametric estimate of the posterior density of the variance parameters σ^2_η and σ^2_κ (top right panel); the modes are not far from the maximum likelihood estimates. The bottom left panel shows the M draws $(\phi_1^{(i)}, \phi_2^{(i)})$ from the posterior of the AR parameter distribution. The triangle delimits the stationary region of the parameter space; the posterior means are 1.48 for ϕ_1 and -0.57 for ϕ_2. Finally, the last panel shows the posterior distribution of the change effect, $-\theta_{\psi 1}$, and the level effect $\theta_\psi(1)$. The 95% confidence interval for the latter is (-0.01, 0.05), which confirms that the output gap has only transitory effects on inflation. The posterior mean of ψ_t does not differ from the point estimates arising from the classical analysis. However, the classical confidence intervals in figure 6 are constructed by replacing $\boldsymbol{\Xi}$ with the ML estimates and thus do not take into account parameter uncertainty (see also section 4.2). It cannot be maintained that the classical estimates are more reliable.

For the GM model, the parameter set $\boldsymbol{\Xi}$ is such that the trend and cycle disturbance variances are replaced by the variances in the two regimes, $\sigma^2_{\eta a}, \sigma^2_{\eta b}, \sigma^2_{\kappa a}, \sigma^2_{\kappa b}$, and under the Markov switching specification (ii), according to which S_t is a first order Markov Chain, includes the transition probabilities $\mathcal{T}_{11}, \mathcal{T}_{00}$.

The steps the GS algorithm need to be amended. An additional step is necessary to draw a sample from the distribution of $\mathbf{S} = (S_0, \ldots, S_n)$ conditional on $\boldsymbol{\alpha}$ and $\boldsymbol{\Xi}$. Notice that this distribution depends on these random vectors only via $\boldsymbol{\eta}, \boldsymbol{\kappa}$, and the elements of Ξ, $\sigma^2_{\eta a}, \sigma^2_{\eta b}, \sigma^2_{\kappa a}, \sigma^2_{\kappa b}, \mathcal{T}_{11}, \mathcal{T}_{00}$. Sampling from the full posterior of the indicator variable $\mathbf{S}$ is achieved by the following algorithm (Carter and Kohn, 1994):

1. Sample $S_n^{(i)}$ from the filtered state probability distribution $P(S_n|\boldsymbol{\alpha}, \boldsymbol{\Xi}, \mathbf{y}) = P(S_n|\boldsymbol{\eta}, \boldsymbol{\kappa}, \boldsymbol{\Xi})$
2. For $t = n-1, \ldots, , 1, 0$, sample $S_t^{(i)}$ from the conditional probability distributional

$$P(S_t|S_{t+1}^{(i)}, \boldsymbol{\eta}, \boldsymbol{\kappa}, \boldsymbol{\Xi}) = \frac{P(S_{t+1}^{(i)}|S_t, \boldsymbol{\Xi})P(S_t|\boldsymbol{\eta}^t, \boldsymbol{\kappa}^t, \boldsymbol{\Xi})}{\sum_{S_t} P(S_{t+1}^{(i)}|S_t, \boldsymbol{\Xi})P(S_t|\boldsymbol{\eta}^t, \boldsymbol{\kappa}^t, \boldsymbol{\Xi})}.$$

   where $\boldsymbol{\eta}^t = (\eta_0, \ldots, \eta_t)$ and $\boldsymbol{\kappa}^t = (\kappa_0, \ldots, \kappa_t)$.

The filtered probabilities, $P(S_t|\boldsymbol{\eta}^t, \boldsymbol{\kappa}^t, \boldsymbol{\Xi})$ are obtained by the following discrete filter:

(i) The filter is started with the initial distribution $P(S_0 = 1|\boldsymbol{\eta}^0, \boldsymbol{\kappa}^0, \boldsymbol{\Xi}) = 1, P(S_0 = 0|\boldsymbol{\eta}^0, \boldsymbol{\kappa}^0, \boldsymbol{\Xi}) = 0$, that is we impose that S_t started in the high volatility regime.

(ii) For $t = 1, 2, \ldots, n$, compute the one-step ahead probability distribution $P(S_t|\boldsymbol{\eta}^{t-1}, \boldsymbol{\kappa}^{t-1}, \boldsymbol{\Xi}) = \sum_{S_{t-1}} P(S_t|S_{t-1}, \boldsymbol{\Xi})P(S_{t-1}|\boldsymbol{\eta}^{t-1}, \boldsymbol{\kappa}^{t-1}, \boldsymbol{\Xi})$.

(iii) Compute the filtered probabilities

$$P(S_t|\boldsymbol{\eta}^t, \boldsymbol{\kappa}^t, \boldsymbol{\Xi}) = \frac{f(\eta_t, \kappa_t|S_t, \boldsymbol{\Xi})P(S_t|\boldsymbol{\eta}^{t-1}, \boldsymbol{\kappa}^{t-1}, \boldsymbol{\Xi})}{\sum_{S_t} f(\eta_t, \kappa_t|S_t, \boldsymbol{\Xi})P(S_t|\boldsymbol{\eta}^{t-1}, \boldsymbol{\kappa}^{t-1}, \boldsymbol{\Xi})},$$

where $f(\eta_t, \kappa_t|S_t, \Xi)$ is the product of two independent Gaussian densities with time-varying scale parameters.

Gerlach *et al.* (2000) have proposed an alternative sampling scheme for the indicator variable S_t which generates samples from $P(S_t|S_{j\neq t}, \mathbf{y}, \boldsymbol{\Xi})$ without conditioning on the states or the disturbances. This is more efficient than the above sampler if S_t is highly correlated with the states or the disturbances, which is not the case in our particular application.

Steps 1 and 2 of the GS algorithm are similar but the full posteriors are understood to be conditional on $\mathbf{S}^{(i-1)}$ as well. Furthermore, an additional step, 2-(g), is added for sampling from the full conditionals of the transition probabilities, $\mathcal{T}_{11}, \mathcal{T}_{22}$, and the steps 2 (b) and 2 (c) are replaced as follows:

(b) Generate $\sigma^{2(i)}_{\eta a}$ and $\sigma^{2(i)}_{\eta b}$ from

$$\sigma^2_{\eta a}|\boldsymbol{\eta}^{(i-1)}, \mathbf{S}^{(i-1)} \sim IG\left(\frac{v_\eta + \sum_t S_t^{(i-1)}}{2}, \frac{\delta_\eta + \sum_t S_t^{(i-1)}\eta_t^{(i-1)^2}}{2}\right),$$

$$\sigma^2_{\eta b}|\boldsymbol{\eta}^{(i-1)}, \mathbf{S}^{(i-1)} \sim IG\left(\frac{v_\eta + \sum_t(1 - S_t^{(i-1)})}{2}, \frac{\delta_\eta + \sum_t(1 - S_t^{(i-1)})\eta_t^{(i-1)^2}}{2}\right).$$

This assumes that the prior distribution is $\sigma^2_{\eta a}$ and $\sigma^2_{\eta b} \sim IG(v_\eta/2, \delta_\eta/2)$.

(c) Generate $\sigma_{\kappa a}^{2(i)}$ and $\sigma_{\kappa b}^{2(i)}$ from

$$\sigma_{\kappa a}^2|\boldsymbol{\kappa}^{(i-1)}, \mathbf{S}^{(i-1)} \sim IG\left(\frac{v_\kappa + \sum_t S_t^{(i-1)}}{2}, \frac{\delta_\kappa + \sum_t S_t^{(i-1)} \kappa_t^{(i-1)^2}}{2}\right),$$

$$\sigma_{\kappa b}^2|\boldsymbol{\kappa}^{(i-1)}, \mathbf{S}^{(i-1)} \sim IG\left(\frac{v_\kappa + \sum_t (1 - S_t^{(i-1)})}{2}, \frac{\delta_\kappa + \sum_t (1 - S_t^{(i-1)}) \kappa_t^{(i-1)^2}}{2}\right).$$

This assumes that the prior distribution is $\sigma_{\kappa a}^2$ and $\sigma_{\kappa b}^2 \sim IG(v_\kappa/2, \delta_\kappa/2)$.

(g) Generate $\mathcal{T}_{11}^{(i)}, \mathcal{T}_{10}^{(i)} = 1 - \mathcal{T}_{11}^{(i)}$ and $\mathcal{T}_{00}^{(i)}, \mathcal{T}_{01}^{(i)} = 1 - \mathcal{T}_{00}^{(i)}$ from the posterior

$$\mathcal{T}_{11}^{(i)}|\mathbf{S}^{(i-1)} \sim B\left(a_1 + N_{11}^{(i-1)}, b_1 + N_{10}^{(i-1)}\right), \mathcal{T}_{00}^{(i)}|\mathbf{S}^{(i-1)} \sim B\left(a_0 + N_{00}^{(i-1)}, b_0 + N_{01}^{(i-1)}\right)$$

where $B(a, b)$ is the Beta distribution, $N_{ij}^{(i-1)}$ is the number of transitions from $S_t^{(i-1)} = i$ to $S_{t+1}^{(i-1)} = j$, and $a_i, b_i, i = 0, 1$ are the parameters of the Beta prior distributions (set equal to $a_1 = b_1 = b_0 = 1, a_0 = 5$). Notice that the transition probabilities are conditionally independent of $\boldsymbol{\alpha}$ and the other elements of $\boldsymbol{\Xi}$, given $\mathbf{S}$.

Figure 8 summarises aspects of the posterior distribution of the cycle, the indicator S_t, and some important parameters using a sample of $M = 2000$ draws from the GS scheme outlined above with a burn-in of 2000 iterations. Interestingly, the output gap interval estimates are more widely dispersed than in the original specification with no Markov switching in the disturbance variances. This is so since the GM specification has a further source of variation and uncertainty, related to the state of Markov Chain S_t, which in turn drives the changes in the volatility regime. As a result the Bayesian interval estimates cannot be compared with the classical ones reported in the bottom left panel of figure 6, since those were derived under the assumption that S_t was deterministic and known, and they make no allowance for parameter uncertainty. The estimated posterior probabilities of being in a high volatility regime confirm the general finding that the main stylised fact is a relatively sharp change point taking place in the first quarter of 1984, although there remains some uncertainty around that date. The nonparametric estimates of the posterior distribution of the transition probabilities $\mathcal{T}_{11}$ and $\mathcal{T}_{00}$ are displayed in the last panel of the figure. The posterior distributions of the variance parameters for the trend and cycle disturbances strongly confirm the great moderation hypothesis, and quantify it further, as both the permanent and transitory disturbances underwent a significant volatility reduction. The posterior means do not differ from the ML estimates reported in table 2: $\mathrm{E}(\sigma_{\eta a}^2|\mathbf{y}) = 0.60, \mathrm{E}(\sigma_{\eta b}^2|\mathbf{y}) = 0.14$ and $\mathrm{E}(\sigma_{\kappa a}^2|\mathbf{y}) = 0.51, \mathrm{E}(\sigma_{\kappa b}^2|\mathbf{y}) = 0.09$. As far as the inflation equation is concerned, the overall conclusion is unchanged: the output gap is a significant source of variation (the value $-\theta_{\psi 1} = 0$ is estimated to be the 2.6 percentile of the posterior distribution of $-\theta_{\psi 1}$, which measures the change effect, but it drives inflation only in the short run, as the null of long run neutrality is accepted (the 95% credible set for $\theta_\psi(1)$ is the interval (-0.01, 0.05)).

3.3 Multivariate extensions

The output gap is related to the deviations of the unemployment rate, u_t, from its "natural rate" or NAIRU via Okun's law. Okun (1962) defined natural unemployment as that level of unemployment occurring when output is equal to its potential, and established an empirical law of strict proportionality between cyclical unemployment and the output gap. Hence, Okun's law is meant to imply that output and the unemployment rate share a common cycle.

Against this background, Clark (1989) estimated a bivariate model of U.S. real output and unemployment such that output and unemployment are decomposed into two unrelated permanent components and the comovements between the two series result from the presence of

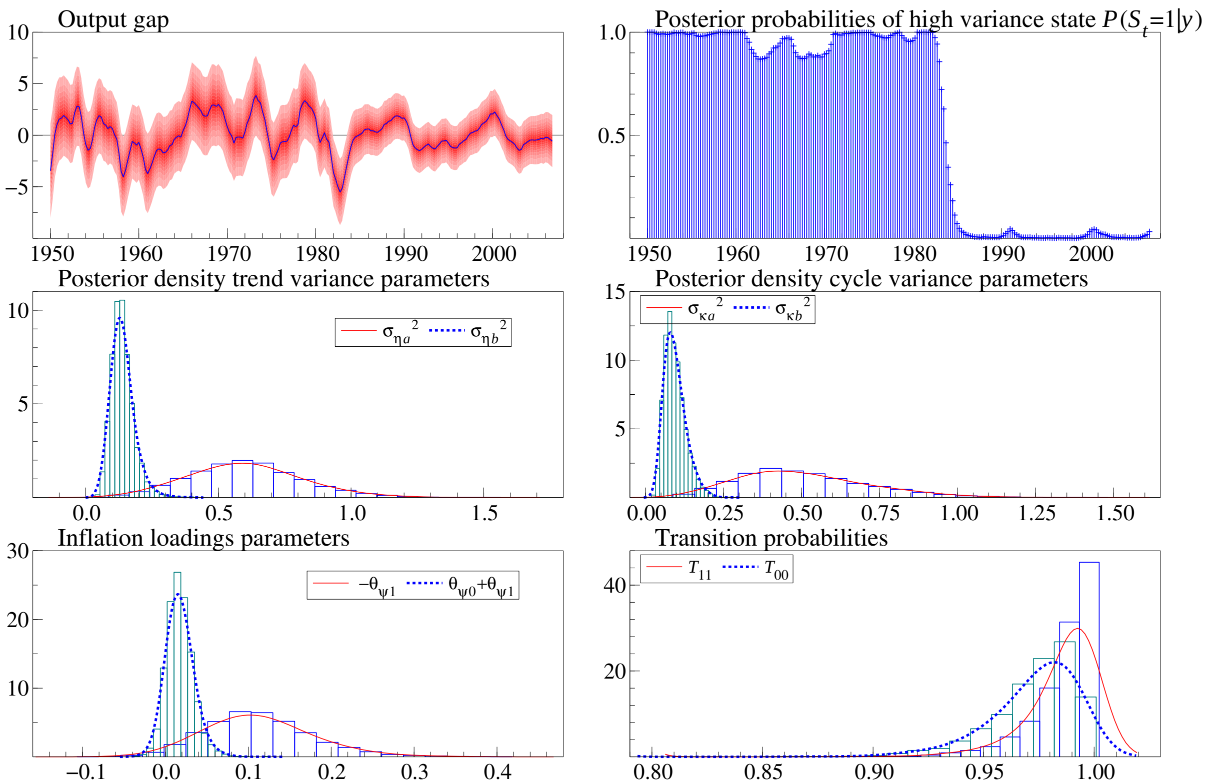

Figure 8: Bayesian estimation of the bivariate output gap model with Markov switching in the variances of the trend and cycle disturbances (GM specification). Point and 95% interval estimates of the output gap; posterior probabilities of the high volatility state, $P(S_t = 1|\mathbf{y})$, and posterior densities of variance and loadings parameters.

a common cycle, represented as an AR(2) stationary component. Apel and Jansson (1999) obtained estimates of the NAIRU and potential output for the U.K, U.S. and Canada, based on an unobserved components model of output, inflation and unemployment rates.

Another important multivariate extension of the basic bivariate model is the *production function approach* (PFA) to the estimation of potential output and the output gap, according to which potential output is obtained from the trend, or "non-inflationary", levels of its structural determinants, such as productivity and factor inputs. This approach is currently one of most popular method of measuring potential output among statistical agencies, being employed by the OECD (2001), the International Monetary Fund (de Masi, 1997), the Congressional Budget Office (2001), and the European Commission (see McMorrow and Roeger 2001).

The PFA assumes that technology can be represented by a Cobb-Douglas production function with constant return to scale on labour, measured by hours worked or by the number of employed persons, and capital:

$$y_t = f_t + \alpha h_t + (1-\alpha)k_t. \tag{23}$$

where f_t is the Solow residual, h_t is hours worked, k_t is the capital stock (all variables expressed in logarithms), and α is the elasticity of output with respect to labour ($0 < \alpha < 1$).

To achieve the decomposition $y_t = \mu_t + \psi_t$, the variables on the right hand side of equation (23) are broken down additively into their permanent (denoted by the superscript P) and transitory (denoted by the superscript T) components, giving:

$$f_t = f_t^{(P)} + f_t^{(T)}, \quad h_t = h_t^{(P)} + h_t^{(T)}, \quad k_t = k_t^{(P)}. \tag{24}$$

It should be noticed that potential capital is always assumed to be equal to its actual value; this is so since capacity utilisation is absorbed in the cyclical component of the Solow residual. Only survey based measures of capacity utilisation for the manufacturing sector are available for the euro area.

Hence potential output is the value corresponding to the permanent values of factor inputs and the Solow residual, while the output gap is a linear combination of the transitory components:

$$\begin{array}{rcl} \mu_t & = & f_t^{(P)} + \alpha h_t^{(P)} + (1-\alpha)k_t, \\ \psi_t & = & f_t^{(T)} + \alpha h_t^{(T)}. \end{array} \tag{25}$$

Hours worked can be separated into four components that are affected differently by the business cycle, as can be seen from the identity $h_t = n_t + pr_t + er_t + hl_t$, where n_t is the logarithm of working age population (i.e., population of age 15-64), pr_t is the logarithm of the labour force participation rate (defined as the ratio of the labour force to the working age population), er_t is the logarithm of the employment rate (defined here as the ratio of employment to the labour force), and hl_t is the logarithm of labour intensity (i.e., average hours worked). Each of these determinants is in turn decomposed into its permanent and transitory component in order to obtain the decomposition:

$$h_t^{(P)} = n_t + pr_t^{(P)} + er_t^{(P)} + hl_t^{(P)}, \quad h_t^{(T)} = pr_t^{(T)} + er_t^{(T)} + hl_t^{(T)}. \tag{26}$$

The idea is that population dynamics are fully permanent, whereas labour force participation, employment and average hours are also cyclical. Moreover, since the employment rate can be restated in terms of the unemployment rate, we can relate the output gap to cyclical unemployment and potential output to structural unemployment. As a matter of fact, since the unemployment rate is one minus the employment rate, $u_t = \log(1 - \exp(er_t))$, the variable $cur_t = -er_t$ (the contribution of the unemployment rate, using a terminology due to Rünstler, 2002), is the first order Taylor approximation to the unemployment rate. Thus, $cur_t^{(P)}$ can be assimilated to the NAIRU and $cur_t^{(T)}$ to the unemployment gap.

The PFA has the appealing feature that it uses a lot of economic information on the determinants of potential output; however, apart from the stringent data requirements (in particular it requires the capital stock and hours worked), it requires the decomposition of the series involved into their permanent and transitory components. Proietti *et al.* (2007) propose a structural time series model-based implementation of the PFA approach, and Proietti and Musso (2007) extend it to carry out a growth accounting analysis for the euro area.

3.4 A multivariate model with mixed frequency data

This section presents the results of fitting a multivariate monthly time series model for the U.S. economy, using quarterly observations for GDP and monthly observations for industrial production, ip_t, the unemployment rate, u_t, and CPI inflation, Δp_t. The equation for the logarithm of GDP is the usual decomposition $y_t = \mu_t + \psi_t$ as in (22), with the important difference that the model is now formulated at the monthly frequency. The CPI equation is also specified as in (22).

Industrial production is included since it is an important timely coincident indicator: the time series model for ip_t is the trend-cycle decomposition $ip_t = \mu_{ip,t} + \theta_{ip}\psi_t + \psi_{ip,t}$, where $\mu_{ip,t} = \mu_{ip,t-1} + \beta_{ip} + \eta_{ip,t}$, and we assume that the trend disturbance is contemporaneously correlated with GDP trend disturbance, η_t, $\eta_{ip,t} \sim \mathrm{N}(0, \sigma^2_{\eta,ip})$, $\mathrm{E}(\eta_t\eta_{ip,t}) = \sigma_{y,ip}$. The cyclical component is the combination of a common cycle and the idiosyncratic cycle $\psi_{ip,t} = \phi_{ip,1}\psi_{ip,t-1} + \phi_{ip,1}\psi_{ip,t-2} + \kappa_{ip,t}$.

The unemployment rate, u_t, is decomposed into the NAIRU, $\mu_{u,t}$, and cyclical unemployment, which is a distributed lag combination of the output gap plus an idiosyncratic component, ψ_{ut}, $u_t = \mu_{u,t} + \theta_{u0}\psi_t + \theta_{u1}\psi_{t-1} + \psi_{ut}$, where the NAIRU is a random walk without drift, $\mu_{u,t} = \mu_{u,t-1} + \eta_{u,t}$, and we assume that $\eta_{u,t} \sim \mathrm{NID}(0, \sigma^2_{\eta,u})$ is independent of any other disturbance in the model, whereas $\psi_{ut} = \phi_{u1}\psi_{u,t-1} + \phi_{u1}\psi_{u,t-2} + \kappa_{ut}$, with $\kappa_{ut} \sim \mathrm{NID}(0, \sigma^2_{\kappa u})$, independently of any other disturbance.

The link between the individual time series equations is provided by the output gap, ψ_t, which acts as the common cycle driving the short run fluctuations; furthermore, the trend disturbances of y_t and ip_t are correlated. As GDP is quarterly, y_t is unobserved, whereas the available observations consist of the aggregated quarterly levels $Y_\tau = \exp(y_{3\tau}) + \exp(y_{3\tau-1}) + \exp(y_{3\tau-2})$, $\tau = 1, 2, \ldots, [n/3]$, where $[\cdot]$ is the integer part of the argument. For the statistical treatment it is useful to convert temporal aggregation into a systematic sampling problem, which is achieved by constructing a cumulator variable, generated by the following time-varying first order autoregression (see Harvey, 1989): $Y^c_t = \varrho_t Y^c_{t-1} + \exp(y_t)$, where ϱ_t is the cumulator coefficient, defined as follows:

$$\varrho_t = \begin{cases} 0 & t = 3(\tau - 1) + 1, \quad \tau = 1, \ldots, [n/3] \\ 1 & \text{otherwise.} \end{cases}$$

Only a systematic sample of the cumulator variable Y^c_τ is available; in particular the end of quarter value is observed, for $t = 3, 6, 9, \ldots, [n/3]$.

The model is represented in the state space form (see Appendix C) with the cumulator variable included in the state vector in the following way. The transition equation $Y^c_t = \varrho_t Y^c_{t-1} + \exp(y_t)$ is nonlinear, but it can be linearised around a trial estimate $\tilde{y}^*_t$ by a first order Taylor series expansion:

$$Y^c_t = \psi_t Y^c_{t-1} + \exp(\tilde{y}^*_t)[1 - \tilde{y}^*_t] + \exp(\tilde{y}^*_t) y_t;$$

replacing $y_t = \mu_t + \psi_t = \mu_{t-1} + \beta + \phi_1\psi_{t-1} + \psi_2\psi_{t-2}$ in the previous expression, Y^c_t can be given a first order inhomogeneous Markovian representation, and thus the model can be cast in state space form, so that conditionally on $\tilde{y}^*_t$ the model is linear and Gaussian.

The fixed interval smoother (C) can be applied to the linearised model to yield an estimate of the components μ_t and ψ_t of the unobserved monthly GDP (on a logarithmic scale) and thus of the monthly series itself. The latter provides a new y_t^* value, which is used to build a new linearised Gaussian model, by a first order Taylor series expansion of Y_t^c. Iterating this process until convergence yields an estimate of the component and of monthly GDP that satisfies the aggregation constraints. See Proietti (2006b) for the theory and applications.

The model was estimated by maximum likelihood using data from January 1950 to December 2006. The estimated parameters for the output gap (standard errors in parentheses) are $\hat{\phi}_1 = 1.73$ (0.021), $\hat{\phi}_2 = -0.744$ (0.037), and $\hat{\sigma}^2_\kappa = 43 \times 10^{-7}$. For potential output $\hat{\beta} = .003$, $\hat{\sigma}^2_\eta = 204 \times 10^{-7}$. The specific cycles for ip_t and u_t are estimated with zero variance, so that the cyclical components of industrial production and unemployment are related to the output gap. The estimated loading of ip_t on ψ_t is $\hat{\theta}_{ip} = 2.454(0.186)$; furthermore, the ip trend disturbances have variance $\hat{\sigma}_{\eta,ip} = 3.74 \times 10^{-7}$, and are positively correlated (with coefficient 0.38) with the GDP trend disturbances. As far as unemployment is concerned, the estimated loadings on ψ_t are $\hat{\theta}_0 = -4.771$ (0.204) and $\hat{\theta}_1 = -2.904$ (0.267); moreover, $\hat{\sigma}^2_{\eta,u} = 9304 \times 10^{-7}$, whereas the irregular disturbance variance was set to zero.

For the inflation equation the output gap loadings are estimated as $\hat{\theta}_{\tau 0} = 0.051$ (0.012) and $\hat{\theta}_{\tau 1} = -0.048$ (0.012); the Wald test for long run neutrality, $H_0 : \theta_{\tau 0} + \theta_{\tau 1} = 0$ takes the value 1.401 with a p-value of 0.24, providing again evidence that the output gap has only transitory effects on inflation. The change effect, $-\theta_{\tau 1}$ is significant and has the expected sign. Finally, the trend disturbance variance for inflation was $\hat{\sigma}^2_\upsilon = 2 \times 10^{-7}$.

Figure 9 presents the smoothed estimates of potential output, the output gap, the NAIRU and core inflation. As a by product, our model produces estimates of monthly GDP that are consistent with the quarterly observed values (the temporal aggregation constraints are satisfied exactly at convergence) and incorporate the information from related series. Comparing the output gap estimates with those arising from the bivariate quarterly model, it can be argued that the use of unemployment series makes a significant difference at the end of the sample. Also, enlarging the information set is beneficial to the reliability of the output gap estimates.

If the model is extended to allow for correlation between the output gap disturbance κ_t and the trend disturbance η_t, as in section 2.5, but in a multivariate set up, the estimated correlation is $\hat{r} = 0.10$ and does not significantly differ from zero. In fact, the model with correlated disturbances has a likelihood of 7263.51, whereas the maximised likelihood of the restricted model ($r = 0$) is 7263.28. Thus, the LR test of $H_0 : r = 0$ takes the value 0.459, with p-value 0.50.

4 The Reliability of the Output Gap Measurement

The reliability of the output gap measurement is the subject of rich debate, and has also strong implications for optimal monetary policy. Orphanides and van Norden (2002) and Camba-Méndez and Rodriguez-Palenzuela (2003) discuss the different sources of uncertainty and their empirical assessment. The former conclude that the real time estimates are unreliable. The conclusion echoes that by Staiger, Stock and Watson (1997) and Laubach (2001) concerning the NAIRU, obtained from a variety of methods. Somewhat different conclusions are reached by Planas and Rossi (2004) and Proietti *et al.* (2007). The implications of the uncertainty surrounding the output gap estimates for monetary policy are considered in Orphanides *et al* (2000) and Ehrmann and Smets (2003), among others.

A full assessment of the output gap reliability is complicated by the very nature of the measurement which, like the NAIRU, core inflation, and so forth, refers to a latent variable, for which there is no underlying “true value” to be elicited by other data collection techniques.

The previous sections have presented different parametric methods that can be used to

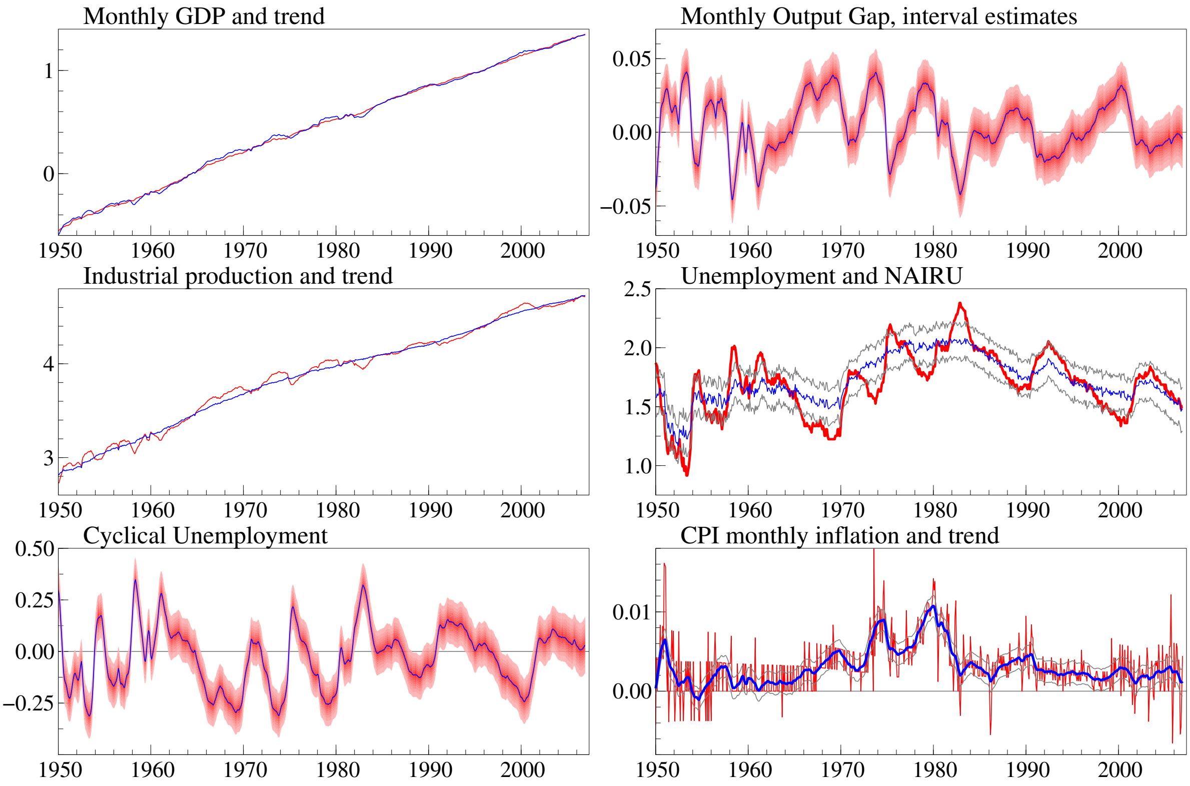

Figure 9: Monthly multivariate output gap model with temporal aggregation constraints. Smoothed estimates of monthly GDP, potential output, output gap, NAIRU, cyclical unemployment and core inflation.

measure the underlying signals. Assume that there is a true output gap ψ_t and that there is an approximating model, denoted by $\mathcal{M}$, providing a representation for it. The model specifies how the observations are related to the object of the measurement. Let us denote by $\psi_{t,\mathcal{M}}$ this (parametric) representation. Now, let $\tilde{\psi}_{t,\mathcal{M}}$ denote the estimator of ψ_t based on model $\mathcal{M}$, i.e., using the representation $\psi_{t,\mathcal{M}}$. We assume that $\tilde{\psi}_{t,\mathcal{M}}$ is the optimal signal extraction method for $\psi_{t,\mathcal{M}}$. How do we judge the reliability of $\tilde{\psi}_{t,\mathcal{M}}$? Reliability is a statement concerning the closeness of $\tilde{\psi}_{t,\mathcal{M}}$ and ψ_t. Following Boumans (2007), two key features are accuracy and precision, as the discrepancy $\tilde{\psi}_{t,\mathcal{M}} - \psi_t$ can be broken down into two components: $(\tilde{\psi}_{t,\mathcal{M}} - \psi_{t,\mathcal{M}}) + (\psi_{t,\mathcal{M}} - \psi_t)$, which are associated respectively to the precision of the method, and to the accuracy or validity of the representation chosen. Given the information set $\mathcal{F}$, precision is measured by (the inverse of) $\mathrm{Var}(\psi_{t,\mathcal{M}}|\mathcal{F}) = \mathrm{E}[(\tilde{\psi}_{t,\mathcal{M}} - \tilde{\psi}_{t,\mathcal{M}})^2|\mathcal{F}]$.

4.1 Validity

Validity is usually difficult to ascertain, as it is related to the appropriateness of $\psi_{t,\mathcal{M}}$ as a model for the signal ψ_t. This is a complex assessment, involving many subjective elements, such as any prior available information and the original motivation for signal extraction. The issue is indissolubly entwined with the nature of ψ_t: the previous paragraphs have considered two main perspectives. The first regards ψ_t as the component of the series that results from the transmission mechanism of demand or nominal shocks. The second view considers ψ_t as the band–pass component of output.

Recently, there has been a surge of interest in model uncertainty and in model averaging. The individual estimates $\tilde{\psi}_{t,\mathcal{M}_i}, i = 1, 2, \ldots, K$, may be combined linearly, giving $\tilde{\psi}_t = \sum_i c_i \tilde{\psi}_{t,\mathcal{M}_i}$, where the coefficients c_i are proportional to the precision the methods, or the posterior probability in a Bayesian setting.

It is more viable to assess two other aspects of validity, namely concurrent and predictive validity. The first is concerned with the contemporaneous relationship between the measure $\tilde{\psi}_{t,\mathcal{M}}$ and a related alternative measure of the same phenomenon. Such measures are rarely available. Although business surveys are implemented with the objective of collecting informed opinions on latent variables, such as the state of the business cycle, they can hardly be considered as providing a measure of the "true" underlying state of the economy.

Predictive validity relates to the ability to forecast future realisations of y_t or related variables; evaluating the mean forecast error yields useful insight on its predictive validity, as possible bias would emerge. This criterion is adopted by a number of authors; for instance, Camba-Mendez and Rodriguez-Palenzuela (2003) and Proietti *et al.* (2007) assess the reliability of alternative output gap estimates through their capability to predict future inflation.

4.2 Precision

A measurement method is precise if repeated measurements of the same quantity are in close agreement. Loosely speaking, precision is inversely related to the uncertainty of an estimate. In the measurement of immaterial constructs the sources of uncertainty would include: (i) *parameter uncertainty*, due to the fact that the core parameters $\boldsymbol{\Xi}$ characterising model $\mathcal{M}$, such as the variance of the disturbances driving the components and the impulse response function, are unknown and have to be estimated; (ii) *estimation error*, the latent components are estimated with a positive variance even if a doubly infinite sample on y_t is available; (iii) *statistical revision*, as new observations become available, the estimate of a signal is updated so as to incorporate the new information; (iv) *data revision.*

The first source can be assessed by various methods in the classical approach; it is automatically incorporated in the interval estimates of the output gap if a Bayesian approach is adopted, as in section 3.2.2. The methods rely on the fundamental result that, under regularity conditions, the maximum likelihood estimator of $\boldsymbol{\Xi}$ has the asymptotic distribution: $\tilde{\boldsymbol{\Xi}} \sim \mathrm{N}(\boldsymbol{\Xi}, \mathbf{V})$, where $\mathbf{V}$ is the inverse of the information matrix. Hamilton (1986) proposed a Bayesian marginalisation approach, which uses $\tilde{\boldsymbol{\Xi}} \sim \mathrm{N}(\boldsymbol{\Xi}, \mathbf{V})$ as a normal approximation to the posterior distribution of Ξ, given the available data. Then, a measure of the uncertainty of the smoothed estimates of the output gap, which embodies parameter uncertainty, is

$$\widehat{\mathrm{Var}}(\psi_t|\mathcal{F}) = \frac{1}{K}\sum_{k=1}^{K} \mathrm{Var}(\psi_t|\mathcal{F}, \tilde{\Xi}^{(k)})] + \frac{1}{K}\sum_{k=1}^{K}\left[\mathrm{E}(\psi_t|\mathcal{F}, \tilde{\Xi}^{(k)}) - \hat{\mathrm{E}}(\psi_t|\mathcal{F})\right]^2, \tag{27}$$

where $\hat{\mathrm{E}}(\psi_t|\mathcal{F}) = \frac{1}{K}\sum_{k=1}^{K}\mathrm{E}(\psi_t|\mathcal{F}, \tilde{\Xi}^{(k)})$, and the $\tilde{\Xi}^{(k)}$s are independent draws from the multivariate normal density $\mathrm{N}(\tilde{\Xi}, \tilde{\mathbf{V}}), k = 1, \ldots, K$, where $\tilde{\mathbf{V}}$ is evaluated at $\tilde{\boldsymbol{\Xi}}$. According to the *delta method* proposed by Ansley and Kohn (1986), expressing the output gap estimates as a linear function of the parameter estimation error $\Xi - \tilde{\Xi}$ gives

$$\widehat{\mathrm{Var}}(\psi_t|\mathcal{F}) = \mathrm{Var}(\psi_t|\mathcal{F}, \tilde{\Xi}) + \mathbf{d}(\tilde{\boldsymbol{\Xi}})'\tilde{\mathbf{V}}\mathbf{d}(\tilde{\boldsymbol{\Xi}}), \qquad \mathbf{d}(\tilde{\boldsymbol{\Xi}}) = \frac{\partial}{\partial \boldsymbol{\Xi}}\mathrm{E}(\psi_t|\mathcal{F}, \tilde{\boldsymbol{\Xi}})\,|_{\boldsymbol{\Xi}=\tilde{\boldsymbol{\Xi}}}, \tag{28}$$

where the derivatives in $\mathbf{d}(\tilde{\boldsymbol{\Xi}})$ are evaluated numerically using the support of the Kalman filter and smoother. Similar methods apply for the real time estimates, with the ML estimator being based on the information set $\mathcal{F}_t$. Quenneville and Singh (2000) evaluate and compare the two methods, and propose enhancements in a Bayesian perspective.

In an unobserved component framework the Kalman filter and smoother provide all the relevant information for assessing (ii) and (iii). For the latter, we can keep track of revisions by using a *fixed-point smoothing* algorithm (see Anderson and Moore, 1979, and de Jong, 1989).

The sources (ii) and (iii) typically arise because the individual components are unobserved and are dependent through time. The availability of additional time series observations helps to improve the estimation of an unobserved component. Multivariate methods are more reliable as they use repeated measures of the same underlying latent variable and this increases the precision of the estimates. It is important to measure the uncertainty that surrounds the real time, or concurrent, estimates, $\mathsf{Var}(\psi_t|\mathcal{F}_t, \tilde{\boldsymbol{\Xi}})$, which are conditional on the information set available to economic agents and policy makers at the time of making the assessment of the state of the economy, as opposed to the historical, or final, estimates. Comparing $\mathrm{Var}(\psi_t|\mathcal{F}_t, \tilde{\boldsymbol{\theta}})$ with the final estimation error variance, $\mathsf{Var}(\psi_t|\mathcal{F}_n, \tilde{\boldsymbol{\Xi}}), n \to \infty$, gives a clue about the magnitude of the revision of the estimates as new observations become available.

In the absence of structural breaks, statistical revisions are sound and a fact of life (i.e., a natural consequence of optimal signal extraction). There is, however, great concern about revisions especially for policy purposes; Orphanides and van Norden (2003) propose temporal consistency as a yardstick for assessing the reliability of output gaps estimates; temporal consistency occurs when real time (filtered) estimates do not differ significantly from the final (smoothed) estimates.

Finally, an additional source of uncertainty is data revision, which concerns y_t. Timely economic data are only provisional and are revised subsequently with the accrual of more complete information. Data revision is particularly relevant for national accounts aggregates, which require integrating statistical information from different sources and balancing it so as to produce internally consistent estimates.

A Linear filters

A linear filter applied to a univariate series y_t is a weighted linear combination of its consecutive values. A time invariant filter can be represented as:

$$\mathrm{w}(L) = \sum_j \mathrm{w}_j L^j. \tag{29}$$

with w_j representing the filter weights. The above filter is symmetric if $\mathrm{w}_j = \mathrm{w}_{-j}$, in which case we can write $\mathrm{w}(L) = w_0 + \sum_j \mathrm{w}_j(L^j + L^{-j})$.

Applying $\mathrm{w}(L)$ to y_t yields $\mathrm{w}(L)y_t$ and has two consequences: the amplitude of the original fluctuations will be compressed or enhanced and the displacement over time of the original fluctuations will be altered. These effects can be fully understood in the frequency domain by considering the frequency response function (FRF) associated with the filter, which is defined as the Fourier transform of (29): $\mathrm{w}(e^{-\imath\omega}) = \sum_j \mathrm{w}_j e^{-\imath\omega j} = \mathrm{w}_{\mathcal{R}}(\omega) + \imath w_{\mathcal{I}}(\omega)$, where $\mathrm{w}_{\mathcal{R}}(\omega) = \sum_j \mathrm{w}_j \cos\omega j$ and $\mathrm{w}_{\mathcal{I}}(\omega) = \sum_j \mathrm{w}_j \sin\omega j$. The last equality stresses that, in general, the FRF is a complex quantity, with $w_{\mathcal{R}}(\omega)$ and $\mathrm{w}_{\mathcal{I}}(\omega)$ representing its real and complex part, respectively. The polar representation of the FRF, $\mathrm{w}(e^{-\imath\omega}) = G(\omega)e^{-\imath Ph(\omega)}$, is written in terms of two crucial quantities, the *gain*, $G(\omega) = |\mathrm{w}(e^{-\imath\omega})| = \sqrt{\mathrm{w}_{\mathcal{R}}(\omega)^2 + \mathrm{w}_{\mathcal{I}}(\omega)^2}$, and the *phase* $Ph(\omega) = \arctan(-\mathrm{w}_{\mathcal{I}}(\omega)/\mathrm{w}_{\mathcal{R}}(\omega))$. The former measures the amplitude effect of the filter, so that if at some frequencies the gain is less than one, then those frequency components will be attenuated in the filtered series; the latter measures the displacement, or the phase shift, of the signal.

If $f_y(\omega)$ denotes the spectrum of y_t, the spectrum of $\mathrm{w}(L)y_t$ is equal to $|\mathrm{w}(e^{-i\omega})|^2 f_y(\omega)$, and therefore the square of the gain function (also known as the *power transfer function*) provides the factor by which the spectrum of the input series is multiplied to obtain that of the filtered series. In the important special case when $\mathrm{w}(L)$ is symmetric, the phase displacement is zero, and the gain is simply $G(\omega) = |\mathrm{w}_0 + 2\sum_{j=1}^m \mathrm{w}_j \cos\omega j|$.

B The Wiener-Kolmogorov filter

The classical Wiener-Kolmogorov prediction theory, which is restricted to stationary processes, deals with optimal signal extraction of an unobserved component. Letting μ_t denote some stationary signal and y_t an indeterministic linear process with Wold representation $y_t = v(L)\xi_t$, $v(L) = 1 + v_1 L + v_2 L^2 + \cdots$, $\sum |v_j| < \infty$, $\xi_t \sim \mathrm{WN}(0, \sigma^2)$, the minimum mean square linear estimator of μ_{t+l} based on a semi-infinite sample $y_{t-j}, j = 0, 1, \ldots, \infty$, is:

$$\tilde{\mu}_{t+l|t} = \frac{1}{\sigma^2 v(L)} \left[\frac{g_{\mu y}(L)}{v(L^{-1})} L^{-l} \right]_+ y_t; \tag{30}$$

here $g_{\mu y}(L)$ denotes the crosscovariance generating function of μ_t and y_t, $g_{\mu y}(L) = \sum_j \gamma_{\mu y,j} L^j$, where $\gamma_{\mu y,j}$ is the crosscovariance at lag j, $\mathrm{E}[(\mu_t - \mathrm{E}(\mu_t))(y_{t-j} - \mathrm{E}(y_t))]$, and for $h(L) = \sum_{j=-\infty}^{\infty} h_j L^j$, $[h(L)]_+ = \sum_{j=0}^{\infty} h_j L^j$, i.e., a polynomial containing only nonnegative powers of L; see Whittle (1983, p. 42). The formula for $l \leq 0$ provides the weights for signal extraction (contemporaneous filtering for $l = 0$).

If an infinite realisation of future y_t was also available, the minimum mean square linear estimator is

$$\tilde{\mu}_{t|\infty} = \frac{g_{\mu y}(L)}{g_y(L)} y_t,$$

where $g_y(L)$ is the autocovariance generating function of y_t, $g_y(L) = |v(L)|^2\sigma^2$, and $|v(L)|^2 = v(L)v(L^{-1})$. If the series is decomposed into two orthogonal components, $y_t = \mu_t + \psi_t$, $g_{\mu y}(L) = g_\mu(L)$ (see Whittle, 1983, ch. 5).

These formulae also hold when y_t and μ_t are nonstationary, see Pierce (1978). As an example of their application, the expressions for the final and concurrent estimators of the trend component for model (2), with $\sigma^2_\eta = 0$ and $\sigma^2_\psi/\sigma^2_\zeta = \lambda$ (Leser-HP filter), are:

$$\tilde{\mu}_{t|\infty} = \frac{1}{1+\lambda|1-L|^4} y_t, \qquad \tilde{\mu}_{t|t} = \frac{\theta(1)}{\theta(L)} y_t,$$

and the corresponding detrending filters are:

$$\tilde{\psi}_{t|\infty} = \frac{\lambda|1-L|^4}{1+\lambda|1-L|^4} y_t, \qquad \tilde{\psi}_{t|t} = \frac{\theta(L)-\theta(1)}{\theta(L)} y_t.$$

Here $\theta(L) = 1+\theta_1 L+\theta_2 L^2$ is the reduced form MA polynomial of the local linear trend model (2). The numerator of the filtered detrended series can be rewritten: $\theta(L) - \theta(1) = \Delta\theta^*(1)L + \Delta^2\theta^*_0$, with $\theta^*(L) = \theta^*_0 + \theta^*_1 L = -(\theta_1+\theta_2) - \theta_2 L$.

The expression for $\tilde{\psi}_{t|\infty}$ is sometimes mistakenly taken to imply that the Leser-HP cycle filter makes stationary series integrated up to the fourth order, due to the presence of $|1-L|^4 = (1-L)^2(1-L^{-1})^2$ in the numerator of the filter. It should be recalled that the above formula holds true only for a doubly infinite sample, and the real time filter for extracting $\tilde{\psi}_{t|t}$ contains only the factor Δ^2.

C State space models and methods

The models considered in this chapter admit the state space representation:

$$\begin{array}{rcll} \mathbf{y}_t &=& \mathbf{Z}_t\boldsymbol{\alpha}_t + \mathbf{G}_t\boldsymbol{\epsilon}_t, & t = 1, 2, \ldots, n, \\ \boldsymbol{\alpha}_t &=& \mathbf{T}_t\boldsymbol{\alpha}_{t-1} + \mathbf{H}_t\boldsymbol{\eta}_t, & \end{array} \tag{31}$$

where $\boldsymbol{\epsilon}_t \sim \mathrm{NID}(\mathbf{0}, \mathbf{I})$, $\boldsymbol{\eta}_t \sim \mathrm{NID}(\mathbf{0}, \mathbf{I})$, and $\mathrm{E}(\boldsymbol{\epsilon}_t\boldsymbol{\eta}'_t) = \mathbf{0}$. The initial conditions are specified as follows: $\boldsymbol{\alpha}_0 = \tilde{\boldsymbol{\alpha}}^*_{0|0} + \mathbf{W}_0\boldsymbol{\delta} + \mathbf{H}_0\boldsymbol{\eta}_0$, so that, $\boldsymbol{\alpha}_1|\boldsymbol{\delta} \sim \mathrm{N}(\tilde{\boldsymbol{\alpha}}^*_{1|0} + \mathbf{W}_1\boldsymbol{\delta}, \mathbf{P}^*_{1|0})$, where $\tilde{\boldsymbol{\alpha}}^*_{1|0} = \mathbf{T}_1\tilde{\boldsymbol{\alpha}}^*_{0|0}, \mathbf{W}_1 = \mathbf{T}_1\mathbf{W}_0, \mathbf{P}^*_{1|0} = \mathbf{H}_1\mathbf{H}'_1 + \mathbf{T}_1\mathbf{H}_0\mathbf{H}'_0\mathbf{T}'_1$. The random vector $\boldsymbol{\delta}$ captures initial conditions for nonstationary state components and it is assumed to have a diffuse distribution, $\boldsymbol{\delta} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_\delta)$, with $\boldsymbol{\Sigma}^{-1}_\delta \to 0$. The matrices $\mathbf{Z}_t, \mathbf{G}_t, t = 1, \ldots, n, \mathbf{T}_t, \mathbf{H}_t, \mathbf{W}_0$ are deterministically related to a set of hyperparameters, $\boldsymbol{\Xi}$.

For instance, for the bivariate model of output and inflation considered in section 3.1, $\mathbf{y}_t$ is a bivariate time series, $\boldsymbol{\alpha}_t = (\mu_t, \beta_t, \psi_t, \psi_{t-1}, \tau_t)'$, $\mathbf{Z}_t = \mathbf{Z} = (\mathbf{z}_y, \mathbf{z}_p)', \mathbf{z}'_y = (1, 0, 1, 0, 0), \mathbf{z}'_p = (0, 0, 0, 0, 1)$, $\boldsymbol{\epsilon}_t = \varepsilon_t/\sigma_\varepsilon$, $\mathbf{G}_t = \mathbf{G} = (0, \sigma_\varepsilon)'$, $\boldsymbol{\eta}_t = (\eta_t/\sigma_\eta, \kappa_t/\sigma_\kappa, \upsilon_t/\sigma_\upsilon)'$, $\boldsymbol{\delta} = (\mu_0, \beta_0, \tau_0)', \tilde{\boldsymbol{\alpha}}^*_{0|0} = \mathbf{0}$,

$$\mathbf{T}_t = \mathbf{T} = \begin{pmatrix} \mathbf{T}_\mu & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_\psi & \mathbf{0} \\ \mathbf{0}' & \mathbf{t}'_p & 1 \end{pmatrix}, \mathbf{T}_\mu = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \mathbf{T}_\psi = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix}, \mathbf{t}_p = \begin{pmatrix} \theta_{\tau 0}\phi_1 + \theta_{\tau 1} \\ \theta_{\tau 0}\phi_2 \end{pmatrix},$$

$$\mathbf{H}_t = \mathbf{H} = \begin{pmatrix} \sigma_\eta & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \sigma_\kappa & 0 \\ 0 & 0 & 0 \\ 0 & \theta_{\tau 0}\sigma_\kappa & 0 \end{pmatrix}, \mathbf{W}_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \boldsymbol{\eta}_0 \sim \mathrm{N}(\mathbf{0}, \mathbf{I}_2), \mathbf{H}_0 = \begin{pmatrix} \mathbf{0} \\ \mathbf{C}_\psi \\ 0 \end{pmatrix},$$

where $\mathbf{C}_\psi$ is such that $\mathrm{E}(\boldsymbol{\psi}_0, \boldsymbol{\psi}'_0) = \mathbf{C}_\psi\mathbf{C}'_\psi$, $\boldsymbol{\psi}_0 = (\psi_0, \psi_{-1})'$.

C.1 The augmented Kalman filter

The Kalman filter (KF) is a fundamental algorithm for the statistical treatment of a state space model. Under the Gaussian assumption it produces the minimum mean square estimator of the state vector along with its mean square error matrix, conditional on past information; this is used to build the one-step-ahead predictor of $\mathbf{y}_t$ and its mean square error matrix. Due to the independence of the one-step-ahead prediction errors, the likelihood can be evaluated via the prediction error decomposition.

The case when $\boldsymbol{\delta}$ is a fixed vector (fixed initial conditions) has been considered by Rosenberg (1973). He showed that $\boldsymbol{\delta}$ can be concentrated out of the likelihood function and that its generalised least square estimate is obtained from the output of an augmented KF. The diffuse case has been dealt with by de Jong (1988).

Defining $\mathbf{A}_{1|0} = -\mathbf{W}_1$, $q_0 = 0, \mathbf{s}_0 = \mathbf{0}, \mathbf{S}_0 = \mathbf{0}$, the augmented KF is given by the following recursive formulae and definitions for $t = 1, \dots, n$:

$$\begin{array}{ll}
\mathbf{v}_t^* = \mathbf{y}_t - \mathbf{Z}_t\tilde{\boldsymbol{\alpha}}^*_{t|t-1}, & \mathbf{V}_t = -\mathbf{Z}_t\mathbf{A}_{t|t-1}, \\
\mathbf{F}_t^* = \mathbf{Z}_t\mathbf{P}^*_{t|t-1}\mathbf{Z}_t' + \mathbf{G}_t\mathbf{G}_t', & \mathbf{K}_t = \mathbf{T}_{t+1}\mathbf{P}^*_{t|t-1}\mathbf{Z}_t'\mathbf{F}_t^{*-1}, \\
q_t = q_{t-1} + \mathbf{v}_t^{*'}\mathbf{F}_t^{*-1}\mathbf{v}_t^*, \;\; \mathbf{s}_t = \mathbf{s}_{t-1} + \mathbf{V}_t'\mathbf{F}_t^{*-1}\mathbf{v}_t^*, & \mathbf{S}_t = \mathbf{S}_{t-1} + \mathbf{V}_t'\mathbf{F}_t^{*-1}\mathbf{V}_t, \\
\tilde{\boldsymbol{\alpha}}^*_{t+1|t} = \mathbf{T}_{t+1}\tilde{\boldsymbol{\alpha}}^*_{t|t-1} + \mathbf{K}_t\mathbf{v}_t^*, & \mathbf{A}_{t+1|t} = \mathbf{T}_{t+1}\mathbf{A}_{t|t-1} + \mathbf{K}_t\mathbf{V}_t \\
\mathbf{P}^*_{t+1|t} = \mathbf{T}_{t+1}\mathbf{P}^*_{t|t-1}\mathbf{T}_{t+1}' + \mathbf{H}_{t+1}\mathbf{H}_{t+1}' - \mathbf{K}_t\mathbf{F}_t^*\mathbf{K}_t' &
\end{array} \tag{32}$$

The diffuse likelihood is defined as follows (de Jong, 1991):

$$\ell(\mathbf{y}_1, \dots, \mathbf{y}_n; \boldsymbol{\Xi}) = -\frac{1}{2}\left(\sum_t \ln|\mathbf{F}_t^*| + \ln|\mathbf{S}_n| + q_n - \mathbf{s}_n'\mathbf{S}_n^{-1}\mathbf{s}_n\right). \tag{33}$$

Denoting $\mathbf{Y}_t = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_t\}$, the innovations, $\mathbf{v}_t = \mathbf{y}_t - \mathrm{E}(\mathbf{y}_t|\mathbf{Y}_{t-1})$, the conditional covariance matrix $\mathbf{F}_t = \mathrm{Var}(\mathbf{y}_t|\mathbf{Y}_{t-1})$, the one-step-ahead prediction of the state vector $\tilde{\boldsymbol{\alpha}}_{t|t-1} = \mathrm{E}(\boldsymbol{\alpha}_t|\mathbf{Y}_{t-1})$, and the corresponding covariance matrices, $\mathrm{Var}(\boldsymbol{\alpha}_t|\mathbf{Y}_{t-1}) = \mathbf{P}_{t|t-1}$, are given by:

$$\begin{array}{lclllcl}
\mathbf{v}_t & = & \mathbf{v}_t^* - \mathbf{V}_t\mathbf{S}_{t-1}^{-1}\mathbf{s}_{t-1}, & & \mathbf{F}_t & = & \mathbf{F}_t^* + \mathbf{V}_t\mathbf{S}_{t-1}^{-1}\mathbf{V}_t', \\
\tilde{\boldsymbol{\alpha}}_{t|t-1} & = & \tilde{\boldsymbol{\alpha}}^*_{t|t-1} - \mathbf{A}_{t|t-1}\mathbf{S}_{t-1}^{-1}\mathbf{s}_{t-1}, & & \mathbf{P}_{t|t-1} & = & \mathbf{P}^*_{t|t-1} + \mathbf{A}_{t|t-1}\mathbf{S}_{t-1}^{-1}\mathbf{A}_{t|t-1}'.
\end{array} \tag{34}$$

C.2 Real time (updated) estimates

The *updated* (or *real time*, *filtered*) estimates of the state vector, $\tilde{\boldsymbol{\alpha}}_{t|t} = \mathrm{E}(\boldsymbol{\alpha}_t|\mathbf{Y}_t)$, and the covariance matrix of the real time estimation error are respectively:

$$\begin{array}{lcl}
\tilde{\boldsymbol{\alpha}}_{t|t} & = & \tilde{\boldsymbol{\alpha}}^*_{t|t-1} - \mathbf{A}_{t|t-1}\mathbf{S}_t^{-1}\mathbf{s}_t + \mathbf{P}^*_{t|t-1}\mathbf{Z}_t'\mathbf{F}_t^{-1}(\mathbf{v}_t^* - \mathbf{V}_t\mathbf{S}_t^{-1}\mathbf{s}_t), \\
\mathbf{P}_{t|t} & = & \mathbf{P}^*_{t|t-1} - \mathbf{P}^*_{t|t-1}\mathbf{Z}_t'\mathbf{F}_t^{*-1}\mathbf{Z}_t\mathbf{P}^*_{t|t-1} + (\mathbf{A}_{t|t-1} + \mathbf{P}^*_{t|t-1}\mathbf{Z}_t'\mathbf{F}_t^{*-1}\mathbf{V}_t)\mathbf{S}_t^{-1}(\mathbf{A}_{t|t-1} + \mathbf{P}^*_{t|t-1}\mathbf{Z}_t'\mathbf{F}_t^{*-1}\mathbf{V}_t)'.
\end{array}$$

C.3 Smoothing

Smoothing deals with the estimation of the components and the disturbances based on the full sample of observations. In the Gaussian case the *fixed interval smoother* provides the minimum mean square estimator of $\boldsymbol{\alpha}_t$ using $\mathbf{Y}_n$, $\tilde{\boldsymbol{\alpha}}_{t|n} = \mathrm{E}(\boldsymbol{\alpha}_t|\mathbf{Y}_n)$, along with its covariance matrix $\mathbf{P}_{t|n} = \mathrm{E}[(\boldsymbol{\alpha}_t - \tilde{\boldsymbol{\alpha}}_{t|n})(\boldsymbol{\alpha}_t - \tilde{\boldsymbol{\alpha}}_{t|n})'|\mathbf{Y}_n]$. The computations can be carried out efficiently using the following backwards recursive formulae, given by Bryson and Ho (1969) and de Jong (1989), starting at $t = n$, with initial values $\mathbf{r}_n = 0, \mathbf{R}_n = \mathbf{0}$ and $\mathbf{N}_n = 0$:

$$
\begin{array}{rcl}
\mathbf{r}_{t-1} & = & \mathbf{L}_t'\mathbf{r}_t + \mathbf{Z}_t'\mathbf{F}_t^{*-1}\mathbf{v}_t, \quad \mathbf{R}_{t-1} = \mathbf{L}_t'\mathbf{R}_t + \mathbf{Z}_t'\mathbf{F}_t^{*-1}\mathbf{V}_t, t = n-1, \ldots, 1. \\
\mathbf{N}_{t-1} & = & \mathbf{L}_t'\mathbf{N}_t\mathbf{L}_t + \mathbf{Z}_t'\mathbf{F}_t^{*-1}\mathbf{Z}_t, \\
\tilde{\boldsymbol{\alpha}}_{t|n} & = & \tilde{\boldsymbol{\alpha}}^*_{t|t-1} - \mathbf{A}_{t|t-1}\mathbf{S}_n^{-1}\mathbf{s}_n + \mathbf{P}^*_{t|t-1}(\mathbf{r}_{t-1} - \mathbf{R}_{t-1}\mathbf{S}_n^{-1}\mathbf{s}_n), \\
\mathbf{P}_{t|n} & = & \mathbf{P}^*_{t|t-1} - \mathbf{P}^*_{t|t-1}\mathbf{N}_{t-1}\mathbf{P}^*_{t|t-1} + (\mathbf{A}_{t|t-1} + \mathbf{P}^*_{t|t-1}\mathbf{R}_{t-1})\mathbf{S}_n^{-1}(\mathbf{A}_{t|t-1} + \mathbf{P}^*_{t|t-1}\mathbf{R}_{t-1})'.
\end{array}
\tag{35}
$$

where $\mathbf{L}_t = \mathbf{T}_{t+1} - \mathbf{K}_t\mathbf{Z}_t'$. A preliminary forward KF pass is required to store the quantities $\tilde{\boldsymbol{\alpha}}^*_{t|t-1}$, $\mathbf{A}_{t|t-1}$, $\mathbf{P}^*_{t|t-1}$, $\mathbf{v}_t^*$, $\mathbf{V}_t$, $\mathbf{F}_t^*$ and $\mathbf{K}_t$.

The smoothed estimates of the disturbances are given by $\mathbf{H}_t\tilde{\boldsymbol{\eta}}_t = \mathrm{E}(\mathbf{H}_t\boldsymbol{\eta}_t|\mathbf{Y}_n) = \mathbf{H}_t\mathbf{H}_t(\mathbf{r}_{t-1} - \mathbf{R}_{t-1}\mathbf{S}_n^{-1}\mathbf{s}_n)$, and $\mathbf{G}_t\tilde{\boldsymbol{\epsilon}}_t = \mathrm{E}(\mathbf{G}_t\boldsymbol{\epsilon}_t|\mathbf{Y}_n) = \mathbf{G}_t\mathbf{G}_t'\left[\mathbf{F}_t^{*-1}(\mathbf{v}_t - \mathbf{V}_t\mathbf{S}_n^{-1}\mathbf{s}_n) + \mathbf{K}_t'(\mathbf{r}_t - \mathbf{R}_t\mathbf{S}_n^{-1}\mathbf{s}_n)\right]$.

C.4 The simulation smoother

The simulation smoother is an algorithm which draws samples from the conditional distribution of the states and the disturbances given the observations and the hyperparameters. Carlin, Polson and Stoffer (1992) proposed a single move state sampler, by which the states are sampled one at a time. This proves to be inefficient in the presence of highly autocorrelated state components. Gamerman (1998) proposed a single move disturbance sampler, which is more efficient since the disturbances driving the components are much less persistent and autocorrelated over time. Along with reparameterization, an effective strategy is blocking, through the adoption of a multimove sampler as in Carter and Kohn (1994) and Früwirth Schnatter (1994), who focus on sampling the states. Again, a more efficient multimove sampler can be constructed by focusing on the disturbances, rather than the states. This is the idea underlying the simulation smoother proposed by de Jong and Shephard (1996).

Let $\boldsymbol{\varsigma}_t = \mathbf{C}[\boldsymbol{\epsilon}_t', \boldsymbol{\eta}_t']'$ denote a subset of the disturbances of the series, with $\mathbf{C}$ being a selection matrix. The structure of the state space model model is such that the states are a (possibly singular) linear transformation of the disturbances and that $\mathbf{G}_t\boldsymbol{\epsilon}_t$ can be recovered from $\mathbf{H}_t\boldsymbol{\eta}_t$ via the measurement equation, which implies that the distribution of $(\boldsymbol{\epsilon}', \boldsymbol{\eta}')'|\mathbf{Y}_n$ is singular. Hence, to achieve efficiency and to avoid degeneracies we need to focus on a suitably selected subset of the disturbances. The simulation smoother hinges on the following factorisation of the joint posterior density:

$$
f(\boldsymbol{\varsigma}_0, \ldots, \boldsymbol{\varsigma}_n|\mathbf{Y}_n) = f(\boldsymbol{\varsigma}_n|\mathbf{y})\prod_{t=0}^{n-1} f(\boldsymbol{\varsigma}_t|\boldsymbol{\varsigma}_{t+1}, \ldots, \boldsymbol{\varsigma}_n; \mathbf{Y}_n).
$$

Conditional random vectors are generated recursively: in the forward step the Kalman filter is run and the innovations, their covariance matrix and the Kalman gain are stored. In the backwards sampling step conditional random vectors are generated recursively from $\boldsymbol{\varsigma}_t|\boldsymbol{\varsigma}_{t+1}, \ldots, \boldsymbol{\varsigma}_n; \mathbf{y}$; the algorithm keeps track of all the changes in the mean and the covariance matrix of these conditional densities. The simulated disturbances are then inserted into the transition equation to obtain a sample from $\boldsymbol{\alpha}|\mathbf{Y}_n$.

A more efficient simulation smoother has been developed by Durbin and Koopman (2002). The gain in efficiency arises from the fact that only the first conditional moments of the states or the disturbances need to be evaluated. Let us redefine $\boldsymbol{\varsigma}_t = (\boldsymbol{\epsilon}_t', \boldsymbol{\eta}_t')'$ and let $\tilde{\boldsymbol{\varsigma}} = \mathrm{E}(\boldsymbol{\varsigma}|\mathbf{Y}_n)$, where $\boldsymbol{\varsigma}$ is the stack of the vectors $\boldsymbol{\varsigma}_t$; $\tilde{\boldsymbol{\varsigma}}$ is computed by the disturbance smoother (see Koopman, 1993, and Appendix C.3). We can write $\boldsymbol{\varsigma} = \tilde{\boldsymbol{\varsigma}} + \boldsymbol{\varsigma}^*$, where $\boldsymbol{\varsigma}^* = \boldsymbol{\varsigma} - \tilde{\boldsymbol{\varsigma}}$ is the disturbance smoothing error, with conditional distribution $\boldsymbol{\varsigma}^*|\mathbf{Y}_n \sim \mathrm{N}(\mathbf{0}, \mathbf{V})$, such that the covariance matrix $\mathbf{V}$ does not depend on the observations, and thus does not vary across the simulations (the diagonal blocks are computed by the smoothing algorithm in Appendix C.3). A sample from $\boldsymbol{\varsigma}^*|\mathbf{Y}_n$ is constructed as follows: we first draw the disturbances from their unconditional Gaussian distribution $\boldsymbol{\varsigma}^+ \sim \text{NID}(0, \mathbf{I})$ and construct the pseudo observations $\mathbf{y}^+$ recursively from $\boldsymbol{\alpha}_t^+ = \mathbf{T}_t\boldsymbol{\alpha}_{t-1}^+ + \mathbf{H}_t\boldsymbol{\eta}_t^+, \mathbf{y}_t^+ = \mathbf{Z}_t\boldsymbol{\alpha}_t^+ + \mathbf{G}_t\boldsymbol{\epsilon}_t^+, t = 1, 2, \ldots, n$, where the initial draw is $\boldsymbol{\alpha}_0^+ \sim \text{N}(\mathbf{0}, \mathbf{H}_0\mathbf{H}_0')$. The Kalman filter and the smoothing algorithm computed on the simulated observations $\mathbf{y}_t^+$ will produce $\tilde{\boldsymbol{\varsigma}}_t^+$, and $\tilde{\boldsymbol{\alpha}}_t^+$, and $\boldsymbol{\varsigma}_t^+ - \tilde{\boldsymbol{\varsigma}}_t^+$ will be the desired draw from $\boldsymbol{\varsigma}^*|\mathbf{Y}_n$. Hence , $\tilde{\boldsymbol{\varsigma}} + \boldsymbol{\varsigma}_t^+ - \tilde{\boldsymbol{\varsigma}}_t^+$ is a sample from $\boldsymbol{\varsigma}|\mathbf{Y}_n \sim \text{N}(\tilde{\boldsymbol{\varsigma}}, \mathbf{V})$.

References

Ansley, C., and Kohn, R. (1986), "Prediction Mean Square Error for State Space Models with Estimated Paramaters", *Biometrika*, 73, 467-473.

Apel, M., and Jansson, P. (1999), "A Theory-Consistent System Approach for Estimating Potential Output and the NAIRU", *Economics Letters*, 64, 271-275.

Artis, M., Marcellino, M. and Proietti, T. (2004). Dating Business Cycles: a Methodological Contribution with an Application to the Euro Area. *Oxford Bulletin of Economics and Statistics*, 66, 537-574.

Basistha, A., and Nelson, C. R. (2007). New measures of the output gap based on the forward-looking new Keynesian Phillips curve. *Journal of Monetary Economics*, 54, 498-511.

Baxter, M., and King, R.G. (1999). Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series. *The Review of Economics and Statistics*, 81, 575–593.

Beveridge, S., and C.R. Nelson (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'Business cycle'. *Journal of Monetary Economics* 7, 151-74.

Brewer, K.R.W. (1979). Seasonal Adjustment of ARIMA Series, *Économie Appliquée*, 1, 7-22.

Boumans, M. (2007). Measurement in Economics: a Handbook. Elsevier.

Bry, G. and Boschan C. (1971). Cyclical Analysis of Time Series: Selected Procedures and Computer Programs, National Bureau of Economic Research, New York.

Bryson, A.E., and Ho, Y.C. (1969). *Applied optimal control: optimization, estimation, and control.* Blaisdell Publishing, Waltham, Mass.

Burns A.F., and Mitchell, W.C. (1946). *Measuring Business Cycles.* New York, NBER.

Camba-Mendez, G. and Rodriguez-Palenzuela, D. (2003). Assessment criteria for output gap estimates. *Economic Modelling*, 20, 529-562

Canova, F. (1998). Detrending and Business Cycle Facts. *Journal of Monetary Economics*, 41, 475-512.

Cappé, O., Moulines, E., and Rydén, T. (2005). *Inference in hidden markov models.* Springer Series in Statistics. Springer, New York.

Carlin, B.P., Polson, N.G. and Stoffer, D.S. (1992). A Monte Carlo Approach to Nonnormal and Nonlinear State Space Modeling. *Journal of the American Statistical Association*, 87, 493500.

Carter, C.K. and Kohn, R. (1994). On Gibbs Sampling for State Space Models. *Biometrika*, 81, 541553.

Chib, S. (2001). Markov Chain Monte Carlo Methods: Computation and Inference. In J.J. Heckman and E. Leamer (eds.), *Handbook of Econometrics: volume 5* North Holland, Amsterdam, 3569-3649.

Christiano, L.J., Fitzgerald, T.J. (2003). The band pass filter. *International Economic Review*, 44, 435–465.

Clark, P. K. (1987), The cyclical component of U.S. economic activity, *The Quarterly Journal of Economics*, 102, 797814.

Clark, P. K. (1989). Trend Reversion in Real Output and Unemployment. *Journal of. Econometrics*, 40, 15-32.

Cox, D.R. (1961). Prediction by exponentially weighted moving averages and related methods. *Journal of the Royal Statistical Society*, Series B, 23, 414-422.

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, 253-278.

Congressional Budget Office (2001), *CBO's Method for Estimating Potential Output: an Update*. CBO Memorandum, Washington, U.S.

de Jong, P. (1989). Smoothing and interpolation with the state space model. *Journal of the American Statistical Association*, 84, 1085-1088.

de Jong, P (1991). The diffuse Kalman filter. *Annals of Statistics* 19, 1073-83.

de Jong, P., and Shephard, N. (1996). The simulation smoother. *Biometrika*, 2, 339-50.

DeMasi, P. (1997). IMF Estimates of Potential Output: Theory and Practice. IMF Working Paper No. 97/177, International Monetary Fund, Washington.

Doménech, R., and Gómez V. (2006). Estimating Potential Output, Core Inflation and the NAIRU as Latent Variables. *Journal of Business and Economic Statistics*, 24, 354-365.

Doornik, J.A. (2006), *Ox: An Object-Oriented Matrix Programming Language*, Timberlake Consultants Press, London.

Durbin, J., and S.J. Koopman (2001). *Time Series Analysis by State Space Methods.* Oxford University Press, Oxford.

Durbin, J., and S.J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89, 603-615.

Ehrmann, M. and Smets, F. (2003). Uncertain potential output: implications for monetary policy. *Journal of Economic Dynamics and Control*, 27, 1611-1638.

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

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models.* Springer Series in Statistics. Springer, New York.

Gamerman, D. (1998). Markov chain Monte Carlo for dynamic generalised linear models. *Biometrika*, 85, 215-227.

Gardner, E.S. (1985). Exponential smoothing: the state of the art. *Journal of Forecasting*, 4, 1-28.

Gardner, E.S. (2006). Exponential smoothing: the state of the art. Part II. *International Journal of Forecasting*, 22, 637-666.

Gerlach, S., and Smets, F. (1999). Output Gaps and Monetary Policy in the EMU Area. *European Economic Review*, 43, 801-812.

Gerlach, R. Carter, C. and Kohn, R. (2000). Efficient Bayesian Inference for Dynamic Mixture Models. *Journal of the American Statistical Association*, 95, 819-828.

Giorno C, Richardson P, Roseveare D, van den Noord P (1995) Estimating Potential Output, Output Gaps and Structural Budget Balances. OECD Working Paper No 152

Gómez, V. (2001). The Use of Butterworth Filters for Trend and Cycle Estimation in Economic Time Series. *Journal of Business and Economic Statistics*, Vol. 19, No 3, 365-373.

Gordon, R.J. (1997), The Time-Varying NAIRU and its Implications for Economic Policy, *Journal of Economic Perspectives*, 11, 11-32.

Hamilton, J.D. (1986). A standard error for the estimated state vector of a state space model. *Journal of Econometrics*, 33, 387-397.

Harvey, A.C. (1989). *Forecasting, Structural Time Series and the Kalman Filter*, Cambridge University Press, Cambridge, UK.

Harvey, A.C. (2001). Testing in unobserved components models, *Journal of Forecasting*, 20, 1-19.

Harvey, A.C., and Jäger, A. (1993). Detrending, stylized facts and the business cycle. *Journal of Applied Econometrics*, 8, 231-247.

Harvey, A. C., and Proietti, T. (2005). *Readings in Unobserved Components Models*. Advanced Texts in Econometrics. Oxford University Press, Oxford, UK.

Harvey, A. C., and Trimbur, T. M. (2003) General model-based filters for extracting trends and cycles in economic time series. *The Review of Economics and Statistics*, 85, 24455.

Harvey, A. C., Trimbur, T. M., and Van Dijk, H. K. (2007). Trends and cycles in economic time series: A Bayesian approach. *Journal of Econometrics*, 140, 618-649.

Hodrick R.J., and Prescott, E.C. (1997). Postwar U.S. Business Cycles: an Empirical Investigation. *Journal of Money, Credit and Banking*, 29, 1-16.

Kaiser, R., and Maravall A. (2005), Combining Filter Design with Model-based Filtering: An Application to Business-cycle Estimation, *International Journal of Forecasting*, 21, 691-710.

King, R.G. and Rebelo, S.T. (1993), Low frequency filtering and real business cycles", *Journal of Economic Dynamics and Control*, 17, 207-231.

Kim, C J., and Nelson, C.R (1999a). Has the U.S. Economy become more Stable? A Bayesian Approach based on a Markov-Switching Model of the Business Cycle". *The Review of Economics and Statistics*, 81, 608-616.

Kim, C. J., and C. Nelson. (1999b). *State Space Models with Regime Switching*. MIT Press, Cambridge, MA.

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

Koopman, S. J. (1993). Disturbance smoother for state space models. *Biometrika*, 80, 117-26.

Kuttner, K.N. (1994). Estimating potential output as a latent variable. *Journal of Business and Economic Statistics*, 12, 361-368.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54, 159-178.

Laubach, T. (2001). Measuring The NAIRU: Evidence From Seven Economies. *The Review of Economics and Statistics*, 83, 218-231.

Leser, C.E.V. (1961)A Simple Method of Trend Construction, *Journal of the Royal Statistical Society B*, 23, 91-107.

McConnell, M.M., and Quiros, G.P. (2000). Output Fluctuations in the United States: What has Changed since the Early 1980s? *American Economic Review*, 90, 14641476.

McMorrow K., and Roeger, W. (2001). Potential Output: Measurement Methods, "New" Economy Influences and Scenarios for 2001-2010. A Comparison of the EU15 and the US. *European Economy - Economic Papers* No. 150. Commission of the European Communities. Directorate-General for Economic and Financial Affairs, Brussels.

Mills, T.C. (2003). *Modelling Trends and Cycles in Economic Time Series*. Palgrave Macmillan, Basingstoke

Morley, J.C., Nelson, C.R., and Zivot, E. (2002). Why are Beveridge-Nelson and Unobserved-Component Decompositions of GDP So Different?. *The Review of Economics and Statistics*, 85, 235-243.

Muth, J.F. (1960). Optimal properties of exponentially weighted forecasts. *Journal of the American Statistical Association*, 55, 299-306.

Nerlove, M.L., Grether, D.M. and Carvalho, J.L. (1995). *Analysis of Economic Time Series: A Synthesis*, Revised edition, Academic Press Inc., New York.

Nyblom J., and Mäkeläinen T. (1983). Comparison of tests for the presence of random walk coefficients in a simple linear model. *Journal of the American Statistical Association*, 78, 856–864.

OECD (2001). *OECD Economic Outlook*, No. 69, June 2001, OECD, Paris.

Okun, A. (1962). Potential GNP: Its Measurement and Significance. Proceedings of the Business and Economic Statistics Section of the American Statistical Association. Reprinted in Okun A. (1970), *The Political Economy of Prosperity* 132-135. Norton, New York

Orphanides A, van Norden S (2002) The Unreliability of Output Gap Estimates in Real Time. *The Review of Economics and Statistics*, 84, 569-583

Orphanides, A., Porter, R., Reifschneider, D., Tetlow, R, and Finan, F. (2000) "Errors in the Measurement of the Output Gap and the Design of Monetary Policy", *Journal of Economics and Business*, 52, 117-143.

Percival, D.B., and Walden, A.T. (1993). *Spectral analysis for physical applications. Multitaper and conventional univariate techniques*, Cambridge University Press, Cambridge, UK.

Pierce, D.A., 1978. Signal Extraction Error in Nonstationary Time Series. *Annals of Statistics* 7: 1303–1320.

Planas, C., and Rossi A. (2004). Can inflation data improve the real-time reliability of output gap estimates? *Journal of Applied Econometrics*, 19, 121-133.

Planas C., Rossi A. and Fiorentini G., (2007), Bayesian analysis of output gap. *Journal of Business and Economic Statistics*, forthcoming.

Proietti T. (2005). Forecasting and Signal Extraction with Misspecified Models. *Journal of Forecasting*, 24, 539-556.

Proietti T. (2004). On the Model Based Interpretation of Filters and the Reliability of Trend-Cycle Estimates. Forthcoming. *Econometric Reviews*.

Proietti T. (2006a), Trend–Cycle Decompositions with Correlated Components. *Econometric Reviews*, 25, 61-84

Proietti T. (2006b). On the Estimation of Nonlinearly Aggregated Mixed Models, *Journal of Computational and Graphical Statistics*, 15, 18-38.

Proietti T., Musso A., and Westermann T. (2007). Estimating Potential Output and the Output Gap for the Euro Area: a Model-Based Production Function Approach. *Empirical Economics*, 33, 85-113.

Proietti T., and Musso A. (2007). Growth accounting for the euro area: a structural approach. ECB Working Paper Series No. 804, European Central Bank, Frankfurt.

Quenneville B, and Singh AC (2000) Bayesian prediction mean squared error for state space models with estimated parameters. *Journal of Time Series Analysis*, 21, 219-236

Ravn, M.O. and Uhlig, H. (2002). On adjusting the HodrickPrescott filter for the frequency of observations. *The Review of Economics and Statistics*, 84, 371376.

Richardson, P., Boone, L., Giorno, C., Meacci, M., Rae, D. and Turner, D. (2000). The Concept, Policy Use and Measurement of Structural Unemployment: Estimating a Time-Varying NAIRU Across 21 OECD Countries. *OECD Working Paper No. 250.*

Rosenberg, B. (1973). Random coefficient models: the analysis of a cross-section of time series by stochastically convergent parameter regression. *Annals of Economic and Social Measurement*, 2, 399-428.

Rudebusch, G., and Svensson, L.E.O. (1998), Policy Rules for Inflation Targeting, in Taylor, J. B., ed., *Monetary Policy Rules*, Chicago University Press, forthcoming.

Rünstler G (2002) The information content of real-time output gap estimates an application to the euro area. ECB Working paper, n 182

Sayed, A.H., and Kailath, T. (2001). A Survey of Spectral Factorization Methods. *Numerical Linear Algebra with Applications*, 8, 467496.

Smets, F. (1998), "Output Gap Uncertainty: Does it Matter for the Taylor Rule?", *BIS Working Paper No. 60*, Bank for International Settlements, Basel, Switzerland.

Staiger, D., Stock, J.H., and Watson M.W. (1997), The NAIRU, Unemployment and Monetary Policy, *Journal of Economic Perspectives*, 11, 33-50.

Stock, J., and M. Watson. 2003. Has the Business Cycle Changed and Why? NBER Macroeconomics Annual 2002 17: 159-218.

Taylor J. B. (1999). A Historical Analysis of Monetary Policy Rules. In J.B. Taylor (ed.), *Monetary Policy Rules.* University of Chicago Press, Chicago, pp. 319-341.

Tiao, G. C., and Xu, D. (1993). Robustness of Maximum Likelihood Estimates for Multi-Step Predictions: The Exponential Smoothing Case. *Biometrika*, 80, 623-641.

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

Whittle P. (1983). *Prediction and Regulation by Linear Least Squares Methods*, Second edition. Basil Blackwell, Oxford.