AUTOMATIC MODEL IDENTIFICATION IN THE PRESENCE OF MISSING OBSERVATIONS AND OUTLIERS *

Víctor Gómez **

D-98009

Abril 1998

** Ministerio de Economía y Hacienda

* This document is an updated version of a previous document written in spanish by the same author titled "Especificación Automática de Modelos ARIMA en Presencia de Observaciones Atípicas", mimeo, Departamento de Estadística e I.O., Universidad Complutense de Madrid, June 1994.

Abstract

Butterworth filters are low–pass filters widely used in electrical engineering. It is shown in the paper that they can be obtained as the solution to a simple signal extraction problem. Details to design and construct Butterworth filters and band–pass filters based on them are given and it is shown how these fixed filters can be applied to economic time series to estimate smooth trends and/or business cycles. If, in addition, a model for the input series is available, a two–step procedure is proposed which circumvents many of the problems of ad–hoc filtering and can significantly improve the quality of the estimated signal.

Keywords: Kalman filter; Signal extraction; ARIMA components model; Smoothing; Butterworth filters; Wiener–Kolmogorov filters.

1 Introduction and Summary

Suppose a time series $z = (z_1, \ldots, z_N)'$ generated by the signal–plus–noise model

$$z_t = s_t + n_t, \tag{1}$$

where s_t is the signal, which in the examples that will interest us will be the trend component, and n_t is a white noise process independent of s_t. We make the following assumptions. The signal s_t follows the ARIMA model $\phi(B)\alpha(B)s_t = \theta_s(B)b_t$, where the polynomial $\alpha(B)$ in the backshift operator, $B^k s_t = s_{t-k}$, has all its roots on the unit circle and degree d, the polynomial $\phi(B)$ has all its roots outside the unit circle and degree p, the polynomial $\theta_s(B)$ has all its roots on or outside the unit circle and degree q_s, and the variables b_t are uncorrelated with the n_t. Also, $\{b_t\}$ and $\{n_t\}$ are serially uncorrelated processes with mean zero, $\mathrm{Var}(b_t) = \sigma_b^2$ and $\mathrm{Var}(n_t) = \sigma_n^2$. The model (1) is nonstationary if $d > 0$.

These assumptions imply that the process $\{z_t\}$ follows the so called reduced form ARIMA model $\phi(B)\alpha(B)z_t = \theta(B)a_t$, where the coefficients in $\theta(B)$ and the variance of a_t are obtained from the equality $\theta(B)a_t = \theta_s(B)b_t + \phi(B)\alpha(B)n_t$. If z_t is nonstationary, we further make assumption A of Bell (1984).

To estimate the signal s_t in (1), three different approaches, which will be labeled A, B and C, can be considered.

A) Cast model (1) into state space form and apply any of the existing algorithms based on the Kalman filter which can handle nonstationary state space models, followed by a corresponding smoothing algorithm. The proposed algorithms are a simple modification of the diffuse Kalman filter of De Jong (1991), properly initialized, and the diffuse fixed point smoother.

B) Make assumption A of Bell (1984) and apply the Wiener–Kolmogorov filter and Tunnicliffe Wilson's algorithm like in Burman (1980).

C) Apply penalized least squares smoothing, which can be described as follows. Let $u_t = \alpha(B)s_t$, $t = d+1, \ldots, N$, $u = (u_{d+1}, \ldots, u_N)'$, $\mathrm{Var}(u)$

$= \Omega$ and $\text{Var}(n_t) = \lambda$ in (1). Then, the problem is minimize

$$\sum_{t=1}^{N}(z_t - s_t)^2 + \lambda u'\Omega^{-1}u.$$

Note that the only approach that allows for the computation of the mean squared errors (MSE) of the estimators is approach A. Using the results of Maravall (1995), approach B can also give the MSE. However, approach A is more flexible and can be easily generalized to the case where, for example, there may be missing observations in model (1) and/or regression variables, where the results of approach B cannot be applied.

Two of the existing approaches to handle nonstationary state space models deal with the problem in all its generality. These are the Diffuse likelihood approach of De Jong (1991) and the marginal likelihood approach of Ansley and Kohn (1985). De Jong (1991) proposed an algorithm, which he called the Diffuse Kalman filter, hereafter referred to as DKF, and Ansley and Kohn (1985) proposed a "modified Kalman filter". This last filter was difficult to implement with existing software and was also conceptually difficult. Recently, Koopman (1997) has proposed an algorithm which is based on the idea of the modified Kalman filter, but is more efficient and a lot simpler to implement. There are also smoothing algorithms corresponding to the approaches of De Jong and Ansley and Kohn, called diffuse smoother and modified smoother.

When the process z_t is stationary, it is well known that applying the Kalman filter and a smoothing algorithm to estimate the signal s_t in (1) is equivalent to first applying the Wiener–Kolmogorov filter to obtain the estimator based on the doubly–infinite sample and then projecting this estimator on the finite sample. This last projection replaces the unkown values in the first estimator with forecasts and backcasts. Bell (1984) proved that, under an assumption which he called assumption A, the Wiener–Kolmogorov filter could also be applied to a complete realization in the nonstationary case. Assumption A of Bell (1984) is a usual one when forecasting with ARIMA models, see Brockwell and Davis (1992), p. 317. In the finite nonstationary situation, the equivalence between Kalman filtering plus smoothing and Wiener–Kolmogorov filtering, applied like in Burman (1980), has been proved by Gómez (1997). The proof was needed because the approach proposed by Burman (1980) lacked a sound theoretical foundation and the results of Bell (1984) were not applicable to finite nonstationary series.

The use of approach C can be traced back to Whittaker (1923), who suggested that the solution of the minimization problem balance a trade–off of goodness of fit to the data and goodness of fit to a smoothness criterion.

A famous example of approach C is the filter proposed by Hodrick and Prescott (1980), hereafter referred to as HP filter, where the particular values $\lambda = 1600$, $\Omega = I$, and $\alpha(B) = \nabla^2$, $\nabla = 1 - B$, are proposed when it is used with quarterly series. It is well known (see, for example, King and Rebelo, 1989), that the HP filter can be given a signal extraction interpretation. That is, the filter is obtained as the filter that corresponds to the estimator of the signal s_t in (1), under the assumption that s_t follows the model $\nabla^2 s_t = b_t$ and $\{b_t\}$ is a white noise sequence with mean zero and variance 1, independent of the n_t, and $\mathrm{Var}(n_t) = 1600$. Since s_t and z_t are nonstationary, under assumption A of Bell (1984), the Wiener–Kolmogorov filter can be applied to a infinite realization of z_t to obtain the minimum mean squared error estimator $\hat{s}_t$ of the signal s_t. The estimator $\hat{s}_t$ is given by an infinite symmetric filter $H_{HP}(B, F)$

$$\hat{s}_t = H_{HP}(B, F) z_t = \nu_0 z_t + \sum_{k=1}^{\infty} \nu_k (B^k + F^k) z_t, \qquad (2)$$

where F is the forward operator, $F^k z_t = z_{t+k}$. The weights ν_t can be obtained from the signal extraction formula

$$H_{HP}(B, F) = 1/(1 + \lambda(1 - B)^2(1 - F)^2). \qquad (3)$$

The question then arises as to whether the finite version of the signal extraction estimator, which, intuitively, is obtained by replacing in (2) the unknown z_t with forecasts and backcasts, can be computed with the approches A and B and if the results of the three approaches coincide.

It is shown in Gómez (1997) that, under the appropriate assumptions in the unobserved components model (1), the three approaches to estimate s_t yield the same result. The algorithmical details are reviewed in appendix A for completeness. These algorithms are implemented in the programs written by the author and used in the applications later in the paper.

The frequency response function $\hat{H}_{HP}(x)$ of the filter $H_{HP}(B, F)$ is obtained by replacing B with e^{-ix} in (3). After some manipulation, we get

$$\hat{H}_{HP}(x) = \frac{1}{1 + \left(\frac{\sin(x/2)}{\sin(x_c/2)}\right)^4}, \tag{4}$$

where x_c is the frequency that corresponds to $\hat{H}_{HP}(x) = 1/2$ and $\lambda = 1/(16\sin^4(x_c/2))$. Since (4) is a real number, it coincides with the gain function of the filter and there is no phase effect. Expression (4) is a special case of the squared gain of a Butterworth filter of the sine version (BFS), which is given by

$$|G(x)|^2 = \frac{1}{1 + \left(\frac{\sin(x/2)}{\sin(x_c/2)}\right)^{2d}}, \tag{5}$$

where $|G(x_c)|^2 = 1/2$. These filters are low–pass filters that depend on two parameters, d and x_c, and, if x_c is fixed, the effect of increasing d is to make the fall sharper. See figure 1. They are autoregressive filters of the form $H(B) = 1/\theta(B)$, where $\theta(B) = \theta_0 + \theta_1 B + \cdots + \theta_d B^d$ and $|G(x)|^2 = H(e^{-ix})H(e^{ix})$. It will be shown later that the denominator in (3) can be factored as $\theta(B)\theta(F)$, where $\theta(B)$ is a polynomial in B of degree 2. We can say then that the filter $H_{HP}(B, F)$ is a "symmetrized" form of a BFS, since $H_{HP}(B, F) = H(B)H(F)$, where $H(B) = 1/\theta(B)$.

The analogy between (4) and (5) suggests that BFS can be given a signal extraction interpretation. This result is proved in the paper for BFS, for Butterworth filters of the tangent version (BFT) and for band–pass filters derived from BFS and BFT. Therefore, for all the symmetrized forms of these filters, which are of the form $H(B)H(F)$, where $H(B)$ is a quotient of polynomials in B, the three approaches, A, B and C, can be used to obtain the filtered series when the filter is applied to a finite series $z = (z_1, \ldots, z_N)'$.

Another question that arises naturally is the following. When symmetrized BFS or BFT are applied and approach B is used, the forecasts and backcasts are obtained with the model implied by the filter and not with an appropiate model fitted to the series. If we could combine both the ARMA structure of the filter and an appropriate model for the series, the performance of the filter would be improved. This topic is investigated in the paper and it is shown that, when using approach B, it is possible to develop a kind of Tunnicliffe Wilson's algorithm for this case. It is also shown that, instead of using a paralell realization in the Tunnicliffe Wilson's type of algorithms, a cascade implementation is possible that, in many cases, is simpler

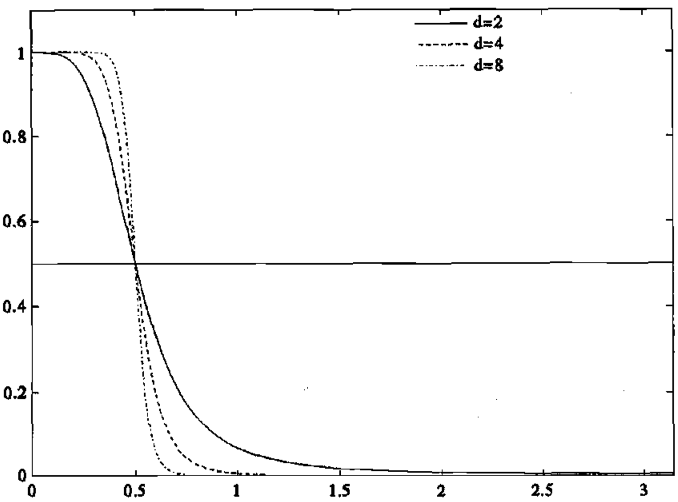

Figure 1: Squared Gain of Butterworth Filters

and more convenient. Fixed ARMA filters can also be applied to components of an unobserved ARIMA components model to obtain smoother trends or business cycle estimates and this is shown with two economic time series.

The structure of the paper is as follows. Butterworth filters are considered in Section 2, where it is shown how their time domain representation can be obtained from the frequency domain formula. It is also proved that the symmetrized forms of these filters can be obtained as the solution to a signal extraction problem. In Section 3, a transformation is proposed that allows for the construction of band–pass filters from low–pass filters. Section 4 gives the details for the design of low–pass and band–pass filters. In Section 5, it is shown how fixed filters should be applied to finite series which are known to follow an ARIMA model and also to components of an unobserved ARIMA components model. Finally, in Section 6, two examples are given of the application of the techniques described in the paper to economic time series.

2 Butterworth Filters

Butterworth filters are low–pass filters widely used in electrical engineering. We consider only the digital forms of these filters, which are of two types. The first one is based on the sine function and will be referred to as BFS, whereas the second one is based on the tangent function and will be referred to as BFT. See Otnes and Enochson (1978).

In the previous section, we saw that a BFS depends on two parameters, d and x_c. This last parameter is the frequency at which the squared gain of the filter is equal to 1/2. By increasing d, the fall of the squared gain function can be made sharper. We also saw in the previous section that the HP filter is a symmetrized form of a BFS which depends only on x_c because d is fixed and equal to 2. Therefore, if we are looking for a low–pass filter that should adapt itself as much as possible to an ideal filter, we find in the symmetrized forms of BFS a much wider class of filters than if we consider the HP filter with a varying x_c. In fact, by choosing d and x_c adequately, we can theoretically approximate an ideal low–pass filter using symmetrized BFS as much as we like.

The squared gain function of a BFT is given by (5) with the sine function replaced with the tangent function. That is, the squared gain function of a BFT is

$$|G(x)|^2 = \frac{1}{1 + \left(\frac{\tan(x/2)}{\tan(x_c/2)}\right)^{2d}}, \tag{6}$$

where, like in the case of a BFS, $|G(x_c)|^2 = 1/2$. The effect of increasing d is also to make the fall sharper, like with BFS.

It is a remarkable fact that symmetrized BFS and BFT can be obtained as best linear estimators, in the mean squared sense, of the signal in models like (1). The decomposition is given by an IMA$(d, 0)$ signal plus white noise for BFS, and by an IMA(d, d) signal plus white noise for BFT; in this last case the MA polynomial is $(1 + B)^d$. Thus, for example, when $d = 1$, the BFS yields the "random walk plus noise" model, and the BFT its canonical version (because there is a spectral zero at frequency $x = \pi$). When $d = 2$, as we saw in the previous section, the BFS yields the HP filter.

The following two theorems, whose proofs are in Appendix B, give the details of this interpretation.

Theorem 1 *Suppose a BFS with parameters d and x_c. Then, the squared gain function of this filter coincides with the gain of the filter obtained by the Wiener–Kolmogorov formula*

$$\hat{s}_t = \frac{1}{1+\lambda(1-B)^d(1-F)^d} z_t \tag{7}$$

to estimate the signal s_t in model (1), where s_t follows the model $\nabla^d s_t = b_t$ and $\lambda = \sigma_n^2/\sigma_b^2$ is given by $\lambda = 1/\left[2^{2d}\sin^{2d}(x_c/2)\right]$. Moreover, the reduced form ARIMA *model for z_t which corresponds to this filter is $\nabla^d z_t = \theta(B)a_t$, where $\theta(B) = \prod_{i=1}^{k}\theta_i(B)$, k is the integer part of $(d+1)/2$, and the $\theta_i(B)$ and the variance σ_a^2 of a_t are given by the formulae*

$$\begin{aligned}
\theta_i(B) &= 1+\theta_{1,i}B+\theta_{2,i}B^2, \quad \theta_{j,i} = \alpha_{j,i}/\alpha_{0,i}, \quad j=1,2 \\
\alpha_{0,i} &= C+\sqrt{D_i}+\sqrt{(C+\sqrt{D_i})^2-1} \\
\alpha_{2,i} &= C+\sqrt{D_i}-\sqrt{(C+\sqrt{D_i})^2-1} \\
\alpha_{1,i} &= 2(C-\sqrt{D_i}), \quad i=1,\ldots,k \\
\sigma_a^2 &= \lambda\prod_{i=1}^{k}\alpha_{0,i}^2,
\end{aligned}$$

where $C = \sin^2(x_c/2)$, $D_i = 1 - 2C\cos((\pi+2(i-1)\pi)/d) + C^2$, and, if d is odd, $\theta_k(B)$ has degree one instead of two and $\alpha_{0,k} = \sqrt{C}+\sqrt{C+1}$, $\alpha_{1,k} = \sqrt{C}-\sqrt{C+1}$. Using the reduced form model, equation (7) can be rewritten as $\hat{s}_t = H(B)H(F)z_t$, where $H(B) = 1/(\sigma_a\theta(B))$.

Theorem 2 *Suppose a BFT with parameters d and x_c. Then, the squared gain function of this filter coincides with the gain of the filter obtained by the Wiener–Kolmogorov formula*

$$\hat{s}_t = \frac{(1+B)^d(1+F)^d}{(1+B)^d(1+F)^d+\lambda(1-B)^d(1-F)^d} z_t \tag{8}$$

to estimate the signal s_t in model (1), where s_t follows the model $\nabla^d s_t = (1+B)^d b_t$ and $\lambda = \sigma_n^2/\sigma_b^2$ is given by $\lambda = 1/\tan^{2d}(x_c/2)$. Moreover, the reduced form ARIMA *model for z_t which corresponds to this filter is $\nabla^d z_t =$*

$\theta(B)a_t$, where $\theta(B) = \prod_{i=1}^{k}\theta_i(B)$, k is the integer part of $(d+1)/2$, and the $\theta_i(B)$ and the variance σ_a^2 of a_t are given by the formulae

$$\begin{aligned}
\theta_i(B) &= 1+\theta_{1,i}B+\theta_{2,i}B^2, \quad \theta_{j,i}=\alpha_{j,i}/\alpha_{0,i}, \quad j=1,2\\
\alpha_{0,i} &= C^2+1+C\sqrt{2(1-D_i)}\\
\alpha_{2,i} &= C^2+1-C\sqrt{2(1-D_i)}\\
\alpha_{1,i} &= 2(C^2-1), \quad i=1,\dots,k\\
\sigma_a^2 &= \lambda\prod_{i=1}^{k}\alpha_{0,i}^2,
\end{aligned}$$

where $C = \tan(x_c/2)$, $D_i = \cos((\pi+2(i-1)\pi)/d)$ and, if d is odd, $\theta_k(B)$ has degree one instead of two and $\alpha_{0,k} = C+1$, $\alpha_{1,k} = C-1$. Using the reduced form model, equation (8) *can be rewritten as $\hat{s}_t = H(B)H(F)z_t$, where $H(B) = (1+B)^d/(\sigma_a\theta(B))$.*

Note that the theoretical signal for a BFT is canonical, because the pseudospectrum of s_t is zero at the π frequency, whereas the one for a BFS is not. This will manifest itself in the fact that BFT will be better approximations to ideal low–pass filters than BFS, although for small values of d the difference will not be substantial.

The two previous theorems show that any of the three approaches considered in this paper can be applied to obtain the finite version of the filtered series $H(B)H(F)z_t$, where $H(B)$ is a BFS or a BFT. Note, therefore, that whenever we apply any of the three approaches to estimate the signal in models implied by Butterworth filters we are using the symmetrized forms of these filters.

EXAMPLE 1 Suppose a yearly univariate series $z = (z_1,\dots,z_N)'$ which follows the model $z_t = s_t + n_t$, where the model for the signal s_t is $\nabla s_t = b_t$, with $\sigma_n^2 = 2$ and $\sigma_b^2 = 1$. Then, the series z_t follows the model $\nabla z_t = (1+\theta B)a_t$ and the parameters θ and $\sigma_a^2 = \mathrm{Var}(a_t)$ can be obtained from theorem 1. Specifically, $d=1$, $\lambda=2$, $C = 1/(4\lambda) = 1/8$, $\alpha_0 = \sqrt{2}$ and $\alpha_1 = -\sqrt{2}/2$. This implies $\sigma_a^2 = 4$ and $\theta = \alpha_1/\alpha_0 = -1/2$.

EXAMPLE 2 Consider the HP filter for quarterly series, which can be obtained by signal extraction from model (1) under the assumption that s_t follows model $\nabla^2 s_t = b_t$, $\sigma_b^2 = 1$ and $\sigma_n^2 = \lambda = 1600$. According to theorem 1, the HP filter is a symmetrized BFS and the reduced form ARIMA model for z_t is $\nabla^2 z_t = \theta(B) a_t$, where $\theta(B) = 1 + \theta_1 B + \theta_2 B^2$, $C = 1/160$, $\alpha_0 = 1.1184$, $\alpha_1 = -1.9875$ and $\alpha_2 = .8941$. This implies $\theta_1 = -1.7771$, $\theta_2 = .7994$ and $\sigma_a^2 = 2001.4$.

3 Band–Pass Filters

Band–pass filters are filters that let pass only those components whose frequencies are in a pre–selected band. These filters can be obtained from low–pass filters by means of a transformation. See Oppenheim and Schaffer (1989), pp. 430–434. Let $[x_{p_1}, x_{p_2}]$, where $x_{p_1} \geq 0$ and $x_{p_1} < x_{p_2} \leq \pi$, be the pass band. Then, a suitable transformation is $z = -s(s-\alpha)/(1-\alpha s)$, where $\alpha = \cos((x_{p_2} + x_{p_1})/2)/\cos((x_{p_2} - x_{p_1})/2)$. Using the inequalities

$$0 < (x_{p_2} - x_{p_1})/2 < (x_{p_2} + x_{p_1})/2 < \pi - (x_{p_2} - x_{p_1})/2$$

and the fact that the cosine function is decreasing in the interval $[0, \pi]$, the inequalities $-1 < \alpha < 1$ hold.

If we apply the previous transformation to a symmetrized BFS, then the following band–pass filter $H_{pbs}(B, F)$ is obtained

$$H_{pbs}(B,F) = \frac{(1-\alpha B)^d(1-\alpha F)^d}{(1-\alpha B)^d(1-\alpha F)^d + \lambda(1-2\alpha B+B^2)^d(1-2\alpha F+F^2)^d},$$

which is the Wiener–Kolmogorov filter to estimate the signal s_t in model (1) when s_t follows the model $(1-2\alpha B+B^2)^d s_t = (1-\alpha B)^d b_t$ and $\lambda = \sigma_n^2/\sigma_b^2$ is like in theorem 1. Since $-1 < \alpha < 1$, we can write $\alpha = \cos\theta$ for a certain $\theta \in [0, \pi]$, so that the polynomial $1 - 2\alpha B + B^2$ has two complex conjugate roots of unit modulus.

If the previous transformation is applied to a symmetrized BFT, then the following band–pass filter $H_{pbt}(B, F)$ is obtained

$$H_{pbt}(B,F) = \frac{(1-B^2)^d(1-F^2)^d}{(1-B^2)^d(1-F^2)^d + \lambda(1-2\alpha B+B^2)^d(1-2\alpha F+F^2)^d},$$

which is the Wiener–Kolmogorov filter to estimate the signal s_t in model (1) when s_t follows the model $(1-2\alpha B+B^2)^d s_t = (1-B^2)^d b_t$ and $\lambda = \sigma_n^2/\sigma_b^2$ is like in theorem 2. Note that in this case the pseudospectrum of the signal is canonical, since it is zero at both the zero and the π frequencies.

The denominators of $H_{pbs}(B, F)$ and $H_{pbt}(B, F)$ can be factored as $\theta^*(B) \times\theta^*(F)\sigma_a^2$, where $\theta^*(B) = \prod_{i=1}^k \theta_i^*(B)$ and k is the integer part of $(d+1)/2$. The polynomials $\theta_i^*(B)$ are obtained from the corresponding $\theta_i(B)$ polynomials of the low–pass filters. For example, if $\theta_i(B) = 1 + \theta_{1,i}B + \theta_{2,i}B^2$, then

$$\begin{aligned}\theta_i^*(B) &= 1+\alpha(\theta_{1,i}-2)B+[\alpha^2(1-\theta_{1,i}+\theta_{2,i})-\theta_{1,i}]B^2\\ &\quad+\alpha(\theta_{1,i}-2\theta_{2,i})B^3+\theta_{2,i}B^4.\end{aligned}$$

4 Design of Low–Pass and Band–Pass Filters

Suppose we want to design a symmetrized BFS and let δ_1, δ_2, x_p and x_s be the specification parameters, so that the gain function $G(x)$, which is the squared of the gain function of the corresponding BFS, should verify $1-\delta_1 < G(x) \le 1$ for $x \in [0, x_p]$ and $0 \le G(x) < \delta_2$ for $x \in [x_s, \pi]$. Since $\sin^2(x/2) = \tan^2(x/2)/(1+\tan^2(x/2))$, we can obtain d and x_c by solving the equations

$$\begin{aligned}
1+\left(\frac{\tan^2(x_p/2)}{1+\tan^2(x_p/2)} \times \frac{1+\tan^2(x_c/2)}{\tan^2(x_c/2)}\right)^d &= \frac{1}{1-\delta_1} \\
1+\left(\frac{\tan^2(x_s/2)}{1+\tan^2(x_s/2)} \times \frac{1+\tan^2(x_c/2)}{\tan^2(x_c/2)}\right)^d &= \frac{1}{\delta_2}.
\end{aligned}$$

First, d is obtained and, if it is not an integer, the nearest integer is chosen. Then, the value of x_c is obtained which corresponds to this integer d in the above equations.

The equations to be solved for the design of a symmetrized BFT are

$$\begin{aligned}
1+\left(\frac{\tan(x_p/2)}{\tan(x_c/2)}\right)^{2d} &= \frac{1}{1-\delta_1} \\
1+\left(\frac{\tan(x_s/2)}{\tan(x_c/2)}\right)^{2d} &= \frac{1}{\delta_2}.
\end{aligned}$$

The way to proceed is like for symmetrized BFS.

If we want to design a symmetrized band–pass filter and the specifications are given by means of the parameters δ_1, δ_2, $x_{p,1}$, $x_{p,2}$, $x_{s,1}$ and $x_{s,2}$, so that the gain function $G(x)$ should verify $1-\delta_1 < G(x) \le 1$ for $x \in [x_{p,1}, x_{p,2}]$ and $0 \le G(x) < \delta_2$ for $x \in [0, x_{s1}]$ and $x \in [x_{s,2}, \pi]$, we may proceed as follows. First, let $x_p = x_{p,2} - x_{p,1}$ and $x_s = x_{s,2} - x_{p,1}$ and design a low–pass filter with the specifications parameters δ_1, δ_2, x_p and x_s. Then, apply the transformation of the previous Section to this low–pass filter to obtain the band–pass filter.

Note that we have not used $x_{s,1}$ in the procedure we have just described to design a band–pass filter. We have implicitly assumed that $x_{s,1}$ is the symmetrical point of $x_{s,2}$ with respect to $(x_{p,1} + x_{p,2})/2$.

5 The use of Fixed Filters Within a Model–Based Approach

Suppose a low–pass filter or a band–pass filter of the type we have considered in the previous Sections, which can be given a signal extraction interpretation, and that we want to apply this filter to a finite series $z = (z_1, \ldots, z_N)'$. Then, we can use any of the three approaches described in this paper to obtain the filtered series. In particular, if approach B is used, the model implied by the filter is used to extend the series with backcasts and forecasts before applying the Wiener–Kolmogorov filter.

If the process $\{z_t\}$ follows an ARIMA model $\phi(B)\alpha(B)z_t = \theta(B)a_t$, which in general will be different from the model implied by the filter, the question naturally arises as to whether the use of this model will improve the performance of the filter at both ends of the series.

Let $\phi^*(B) = \phi(B)\alpha(B)$ and let the filter be $H(B)H(F)$, where $H(B) = \gamma(B)/\beta(B)$. Suppose first that we know a complete realization of the process $\{z_t\}$ and let $\{x_t\}$ be the filtered series $x_t = H(F)H(B)z_t$ and $y_t = H(B)z_t$. Then, y_t follows the model $\phi^*(B)\beta(B)y_t = \theta(B)\gamma(B)a_t$. Since the series z_t also follows the backward model $\phi^*(F)z_t = \theta(F)v_t$, projecting onto the finite sample $z = (z_1, \ldots, z_N)'$, it is obtained that

$$\begin{aligned} \phi^*(B)\beta(B)y_t &= 0 \quad t \geq N + q + a + 1 \\ \phi^*(F)z_t &= 0 \quad t \leq -q, \end{aligned}$$

where q is the degree of $\theta(B)$ and a is the degree of $\gamma(B)$. Then, if p^* is the degree of $\phi^*(B)$ and b is the degree of $\beta(B)$, the following algorithm can be used for the cascade implementation of the filter

1. Solve the system

$$\begin{aligned} \beta(B)y_t &= \gamma(B)z_t \quad t = -q+1, \ldots, p^* - q \\ \phi^*(F)y_t &= 0 \qquad\qquad t = -q-b+1, \ldots, -q \end{aligned}$$

   where $q + a$ backcasts are needed: $\hat{z}_{-q-a+1}, \ldots, \hat{z}_0$.
   For $t = p^* - q + 1, \ldots, N + q + 2a$, obtain y_t from the recursion $\beta(B)y_t = \gamma(B)z_t$, where $q + 2a$ forecasts are needed: $\hat{z}_{N+1}, \ldots, \hat{z}_{N+q+2a}$.

2. Solve the system

$$\begin{aligned} \beta(F)x_t &= \gamma(F)y_t \quad t = N+q+a-b-p^*+1,\ldots,N+q+a \\ \phi^*(B)\beta(B)x_t &= 0 \qquad\qquad t = N+q+a+1,\ldots,N+q+a+b \end{aligned}$$

For $t = N+q+a-b-p^*,\ldots,1$, obtain x_t from the recursion $\beta(F)s_t = \gamma(F)y_t$.

Note that if $\beta(B) = 1$, the filter is a symmetric moving average, which is factored as $\gamma(F)\gamma(B)$. In this case, it is not necessary to solve the two systems. All that is needed is to generate first y_t from $y_t = \gamma(B)z_t$ and then x_t from $x_t = \gamma(F)y_t$. For that, a backcasts and a forecasts are required. Thus, the previous algorithm can be considered as a generalization to ARMA filters of the procedure used by the program X11 ARIMA for finite moving average filters. See Dagum (1980).

Consider the unobserved ARIMA components model $z_t = p_t + s_t + w_t$, where p_t is the trend, s_t is the seasonal and w_t is the irregular component, and suppose that approach B is used to estimate the components. Then, it may be the case that we are interested in estimating a smoother trend component $p_{1,t}$ than p_t. This could be achieved by setting $p_{1,t} = H(B)H(F)p_t$ and $p_t = p_{1,t} + p_{2,t}$, where $H(B)H(F)$ is an appropriate low–pass filter and $p_{2,t} = (1-H(B)H(F))p_t$. Suppose first that $p = (p_1,\ldots,p_N)'$ is known (although it is not observed). Then, we could apply the previous cascade implementation to obtain the minimum mean squared error estimator of $p_{1,t}$ based on p. The model for the series that we would use would be the model of p_t. Now, since p is not observed, we have to project it onto the observed series $z = (z_1,\ldots,z_N)'$ in order to obtain the estimator $\hat{p}_{1,t}$ of $p_{1,t}$. If $\hat{p}_t$ is the minimum mean squared error estimator of p_t based on z in $z_t = p_t + s_t + w_t$, then $\hat{p}_{1,t}$ is obtained by replacing p_t in the equations for the cascade implementation with $\hat{p}_t$. This algorithm constitutes a simple way to calculate $\hat{p}_{1,t} = H(B)H(F)\hat{p}_t$.

The previous result suggests a sensible procedure to estimate smooth trends or business cycles in two steps. In the first step, apply a model–based procedure to obtain a canonical decomposition $z_t = p_t + s_t + w_t$. Then, apply a symmetrized BFS, BFT, or band–pass filter of Section 3, to the trend estimator $\hat{p}_t$ obtained in the first step using the cascade implementation described in the previous paragraph. The model to use for p_t would be the model for that component given by the canonical decomposition. The business cycle, for example, would be the filtered series, where the filter is an appropiate band–pass filter. This procedure has the advantage that if, for example, the input series is white noise, the first step of the procedure would detect it and $\hat{p}_t$ would be zero. So the procedure automatically safeguards the user against a bad use of the fixed filter. Also, since the components given by the first step are "canonical", the fixed filter would extract all existing power in the frequency band of interest, free of noise because all noise goes to the irregular. Two examples of this procedure will be given in the next Section.

6 Applications

Several programs in Fortran have been written by the author to implement the methodology outlined in the paper. Specifically, these programs include programs to design symmetrized BFS and BFT and band-pass filters obtained from them, and programs to apply the three approaches. In case of approach B, there is also a program to implement the procedure of the previous Section, where a model for the series is known. These programs are available from the author upon request.

Programs TRAMO and SEATS of Gómez and Maravall (1996), have been used for automatic model identification (including a test for the logarithmic transformation) and model estimation (TRAMO), and signal extraction (SEATS), based on the canonical decomposition of the reduced form ARIMA model for the series. These programs are available at the Internet address

http://www.bde.es

To simplify the exposition, the automatic outlier detection and correction facility of TRAMO has not been used, but this aspect could be easily incorporated into the proposed procedure.

The first example is the series of quarterly US GNP, from the first quarter of 1951 until the fourth quarter of 1985. The series can be taken from Citibase data bank. Using TRAMO, the multiplicative ARIMA model $(0,1,1)(0,1,1)_4$ is specified for the logs of the data and the model parameters are estimated. The fit is acceptable, although the residuals show some departure from normality. This is due to the presence of two outliers (transitory changes), at 1984-I and 1958-I. But, as mentioned above, we do not correct for the effect of these outliers and the model is accepted. After having passed the model

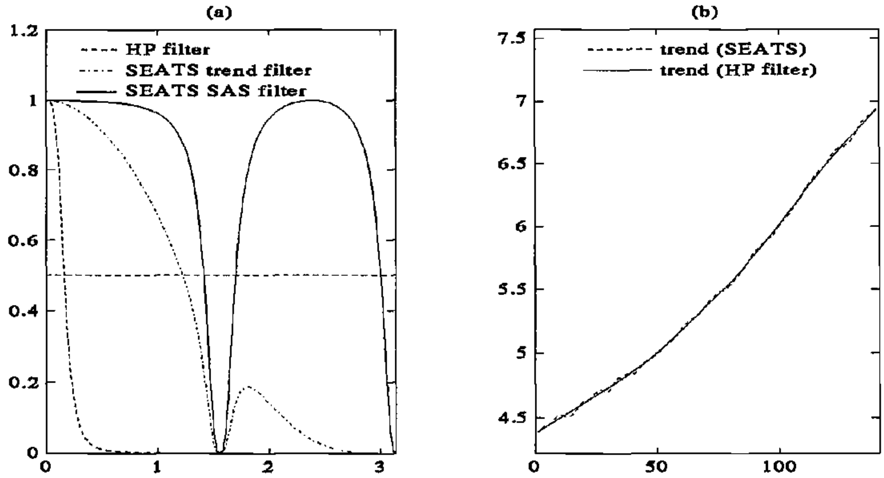

Figure 2: (a) Gain Functions of HP filter and SEATS Trend and SAS Filters. (b) Trends estimated by SEATS and HP filter.

and the parameter estimates to SEATS, signal extraction is performed. Then, the HP filter, which, as shown previously, is a symmetrized BFS, is applied to the seasonally adjusted series (SAS). It is shown in Gómez (1997) for this example that using the three approaches, A, B and C, to estimate the signal s_t in (1) leads to practically the same results.

The gain functions of the HP filter and the trend and SAS filters used by SEATS are displayed in figure 2(a), whereas, in figure 2(b), one can see the trend component estimated by SEATS and the smoother trend obtained by filtering the SAS with the HP filter. The HP filter is a low–pass filter that approximates rather well an ideal filter that passes all components with periods greater than thirty two quarters (eight years).

The cycle estimated by the HP filter is the difference between the SAS and the trend estimated with this filter. This cycle contains however many components with periods between two and six quarters, which should not be properly considered as part of a cyclical component. This is due to the fact that these components are in the seasonally adjusted series and are not removed by the HP filter. See figure 2(a).

Suppose that, as in Burns and Mitchell (1946), we define business cycles as those cyclical components of no less than six quarters (eighteen months) and no more than thirty two quarters (eight years) in duration. Then, it would be interesting to design a band–pass filter corresponding to this definition, apply this filter to the trend estimated with SEATS, using the model for the trend given by the canonical decomposition, and compare the estimated cycle with the cycle given by the HP filter. To this end, we first design a band–pass filter, based on a BFT, with the specifications $\delta_1 = .1$, $\delta_2 = .1$, $x_{p,1} = .0625\pi$, $x_{p,2} = .3\pi$ and $x_{s,2} = .4\pi$, where $x_{p,1}$ corresponds to a period of 32 quarters (8 years) and $x_{p,2}$ to 6.67 quarters (a little more than a year and a half). The parameters d and x_c for the symmetrized BFT, obtained by solving the equations of Section 4, are $d = 5$ and $x_c = .9073$. The gain functions of the band–pass filter so designed and the trend function given by SEATS are displayed in figure 3(a). The trend component p_t given by the canonical decomposition in SEATS follows the ARIMA model $(0,2,2)$, with moving average polynomial $1+0.0877B-0.9123B^2$, where B is the backshift operator. The cycle obtained by applying the band–pass filter to the trend estimator $\hat{p}_t$ given by SEATS, using the model for the theoretical component in the manner described in the previous Section, can be seen in figure 3(b). Also in this figure, we can see the cycle obtained by the HP filter. It is clear that this last cycle is more volatile than the one obtained with the band–pass filter.

The second example is the series of monthly airline passengers of Box and Jenkins (1976), from January 1949 until December 1960. Using TRAMO again, the multiplicative ARIMA model $(0,1,1)(0,1,1)_{12}$ is specified for the logs of the data, the model parameters are estimated, and the fit is acceptable. The model and the parameter estimates are passed on to SEATS and signal extraction is performed.

Suppose that one is interested in obtaining a smoother trend than the trend given by SEATS since, by inspecting figure 4(a), which displays the trend filter used by SEATS, it is easy to see that this last filter is not a very restrictive low–pass filter. We design, for example, a symmetrized BFS that approximates an ideal low–pass filter that passes all components with a period greater than eight years. This can be achieved with the specification $\delta_1 = .1$, $\delta_2 = .01$, $x_p = .02\pi$, and $x_s = .05\pi$, since the frequency that corresponds to a period of 96 months (8 years) is $.02083\pi$. Solving the equations of Section 4, the parameters $d = 4$ and $x_c = .0827$ are obtained. The gain function of this symmetrized BFS can be seen in figure 4(a). We now apply the symmetrized

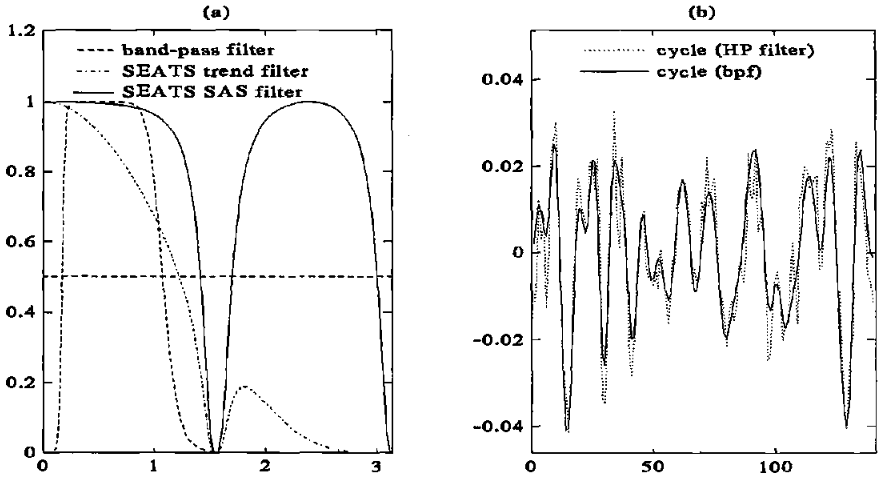

Figure 3: (a) Gain Functions of band–pass, and SEATS Trend and SAS Filters. (b) Cycles estimated by band–pass filter and HP filter.

BFS to the trend estimator $\hat{p}_t$ given by SEATS, using the model for the theoretical component p_t in the manner described in the previous Section. The model that follows p_t is $(0,2,2)$, with moving average polynomial $1 + 0.0478B - 0.9522B^2$, where B is the backshift operator. The trend estimated using the symmetrized BFS and the trend estimated by SEATS are displayed in figure 4(b). In order to estimate a cyclical component, using the definition of the business cycle of Burns and Mitchell (1946), we design a band–pass filter, based on a symmetrized BFT, with the specifications $\delta_1 = .1$, $\delta_2 = .01$, $x_{p,1} = .02\pi$, $x_{p,2} = .08\pi$ and $x_{s,2} = .15\pi$, where $x_{p,1}$ corresponds to a period of 100 months (8.33 years) and $x_{p,2}$ to 25 months (a little more than two years). The parameters d and x_c for the symmetrized BFT, obtained by solving the equations of Section 4, are $d = 4$ and $x_c = .2475$. The gain functions of the band–pass filter so designed and the trend function given by SEATS are displayed in figure 5(a). The cycle obtained by applying the band–pass filter to the trend estimator $\hat{p}_t$ given by SEATS, using the model for the theoretical component p_t in the manner described in the previous Section, is displayed in figure 5(b).

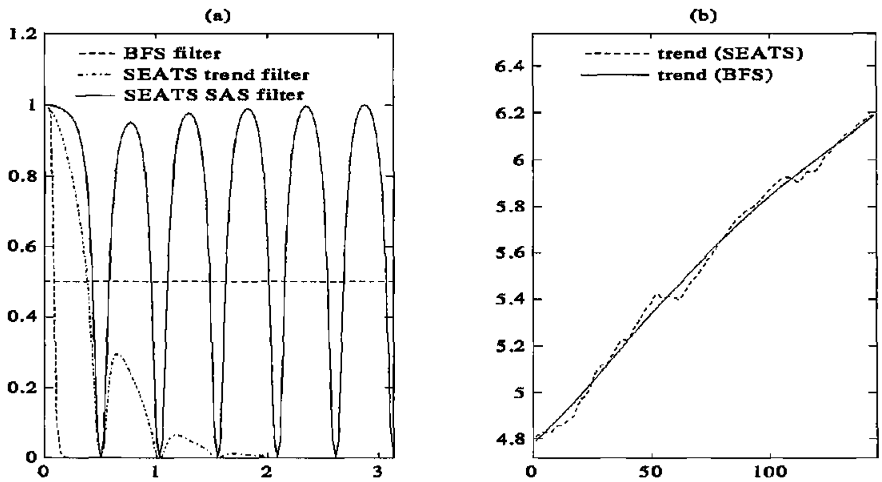

Figure 4: (a) Gain Functions of symmetrized BFS and SEATS Trend and SAS Filters. (b) Trends estimated by SEATS and symmetrized BFS.

APPENDIX A: ALGORITHMICAL DETAILS

Details of Approach A

Among the state space representations of ARIMA models, we select that of Gómez and Maravall (1994), which is an extension to nonstationary series of the representation originally proposed by Akaike (1974) for ARMA models.

Letting $r = max\{p+d, q_s+1\}$, $\phi^*(B) = \phi(B)\alpha(B)$ and defining $\phi_i^* = 0$ when $i > p+d$, the state space representation for model (1) is given by

$$z_t = H'x_t + n_t \qquad \text{(A.1)}$$

$$x_{t+1} = Fx_t + Gb_{t+1}, \qquad \text{(A.2)}$$

where

$$F = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -\phi_r^* & -\phi_{r-1}^* & -\phi_{r-2}^* & \dots & -\phi_1^* \end{bmatrix}, \qquad \text{(A.3)}$$

$x_t = (s_t, s_{t+1,t}, \dots, s_{t+r-1,t})'$, $H = (1, 0, \dots, 0)'$, $G = (1, \psi_1^*, \dots \psi_{r-1}^*)'$ and the ψ_i^* weights are the coefficients obtained from $\psi^*(B) = \theta_s(B)/\phi^*(B) = \sum_{i=0}^{\infty} \psi_i^* B^i$. The

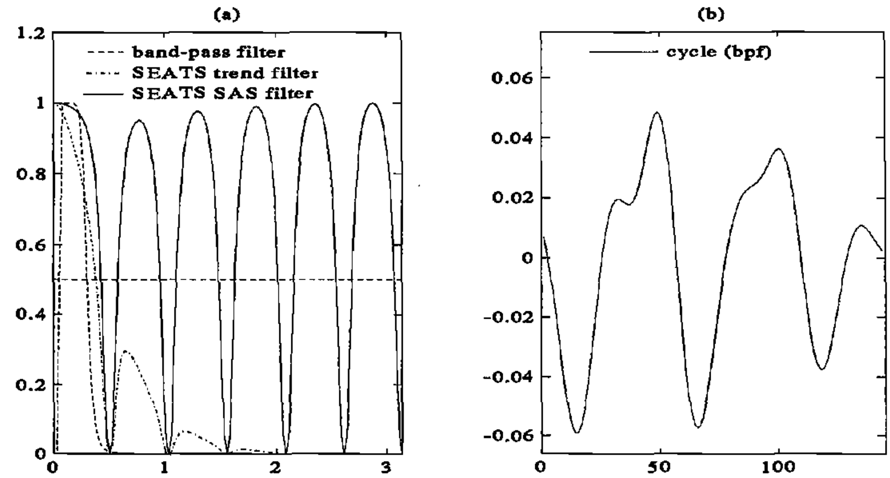

Figure 5: (a) Gain Functions of band–pass, and SEATS Trend and SAS Filters. (b) Cycle estimated by band–pass filter.

elements of the state vector are defined as $s_{t+i,t} = s_{t+i} - \psi_0^* b_{t+i} - \cdots - \psi_{i-1}^* b_{t+1}$, $i = 1, \ldots, r-1$. They are the predictors of s_{t+i} based on the semi-infinite sample $\{s_j : j \leq t\}$.

Since the process $\{s_t\}$ follows an ARIMA model, proceeding like in Bell (1984), it can be generated as linear combinations of some starting values and elements of the stationary (differenced) process $u_t = \alpha(B)s_t$. Let the starting values be $\delta = (s_{1-d}, \ldots, s_0)'$. Then, following Bell (1984), the s_t can be generated from $s_t = A_t'\delta + \sum_{i=0}^{t-1} \xi_i u_{t-i}$, where $t > 0$, $\xi(B) = \sum_{i=0}^{\infty} \xi_i B^i$ and the $A_t = (A_{1t}, \ldots, A_{dt})'$ can be recursively obtained.

Like in Gómez and Maravall (1994), p. 615, it can be shown that the initial state vector x_1 verifies $x_1 = A\delta + \Xi U$, where $A = [A_1, \ldots, A_r]'$, Ξ is the lower triangular matrix with rows the vectors $(\xi_{j-1}, \xi_{j-2}, \ldots, 1, 0, \ldots, 0)$, $j = 1, \ldots, r$, $U = (u_1, u_{2,1}, \ldots, u_{r,1})'$ and $u_{i,1} = \mathrm{E}(u_i | u_t : t \leq 1)$, $i > 1$.

In the previous expression for x_1, δ models uncertainty with respect to the initial conditions and its distribution is unknown. Therefore, the ordinary Kalman filter cannot be applied and some device has to be used to handle δ, which can be considered as a vector of nuisance random variables.

For algorithmical purposes, we will use the approach of De Jong (1991) in this paper. Using the transition equation (A.2), if $z = (z_1, \ldots, z_N)'$ is the observed series, we have the following representation

$$z = X\delta + \epsilon, \tag{A.4}$$

where, partitioning $X = (X_1, \ldots, X_N)'$ and $\epsilon = (\epsilon_1, \ldots, \epsilon_N)'$ conforming to $z = (z_1, \ldots, z_N)'$, the X_t' and ϵ_t can be recursively obtained, the distribution of ϵ is known, $\mathrm{E}(\epsilon) = 0$, and $\mathrm{Cov}(\delta, \epsilon) = 0$.

Let $\mathrm{Var}(b_t, n_t)' = \sigma_b^2 \mathrm{diag}(1, \lambda)$, where $\lambda = \sigma_n^2/\sigma_b^2$, and $\mathrm{Var}(\epsilon) = \sigma_b^2 \Sigma$ in (A.4). Following De Jong (1991), suppose that δ is independent of the $\{b_t\}$ and $\{n_t\}$, has mean 0 and covariance matrix $\sigma_b^2 C$, and take the limit $C^{-1} \to 0$ to make it diffuse. Assuming normality in n_t, b_t and δ and letting $l(z)$ be the log–likelihood of z in (A.4) it is shown in De Jong (1991) that, apart from a constant, as $C^{-1} \to 0$,

$$\begin{aligned} l(z) + \frac{1}{2}\ln|\sigma_b^2 C| \to \; & -\frac{1}{2}\{(N-d)\ln(\sigma_b^2) + \ln|\Sigma| + \ln|X'\Sigma^{-1}X| \\ & + (z - X\hat{\delta})'\Sigma^{-1}(z - X\hat{\delta})/\sigma_b^2\}, \end{aligned} \tag{A.5}$$

where $\hat{\delta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}z$ and the mean squared error (Mse) of $\hat{\delta}$ is $\mathrm{Mse}(\hat{\delta}) = \sigma_b^2(X'\Sigma^{-1}X)^{-1}$. The limit expression in (A.5) is the diffuse log–likelihood. The parameter σ_b^2 can be concentrated out of the diffuse log–likelihood by replacing σ_b^2 in (A.5) with its maximum likelihood estimator $\hat{\sigma}_b^2 = (z - X\hat{\delta})'\Sigma^{-1}(z - X\hat{\delta})/(N-d)$.

Thus, making δ diffuse implies that (A.4) can be considered as a generalized linear regression model (GLS), where δ is the vector of regression parameters and $\hat{\delta}$ and $\hat{\sigma}_b^2$ are the GLS estimators.

In order to evaluate the diffuse log–likelihood efficiently, let $\Sigma = LL'$, with L lower triangular, be the Cholesky decomposition of $\Sigma = \mathrm{Var}(\epsilon)/\sigma_b^2$ and suppose that an efficient algorithm exists to compute $L^{-1}z$, $L^{-1}X$ and $|L|$. This algorithm is a slight modification of the DKF, which will be described later. Then, premultiplying (A.4) by L^{-1}, it is obtained that

$$L^{-1}z = L^{-1}X\delta + L^{-1}\epsilon, \tag{A.6}$$

where $\mathrm{Var}(L^{-1}\epsilon) = \sigma_b^2 I_N$. Therefore, model (A.6) is an ordinary linear regression model. The GLS estimators $\hat{\delta}$ and $\hat{\sigma}_b^2$ can now be efficiently and accurately obtained using the QR algorithm, as suggested by Kohn and Ansley (1985). This last algorithm premultiplies both $L^{-1}z$ and $L^{-1}X$ by an orthogonal matrix Q to obtain $v = QL^{-1}z$ and $(U', 0')' = QL^{-1}X$, where U is a nonsingular $d \times d$ upper triangular matrix. Then, $\hat{\delta} = U^{-1}v_1$ and $\hat{\sigma}_b^2 = v_2'v_2/(N-d)$, where $v = (v_1', v_2')'$, v_1 has dimension d and v_2 has dimension $N - d$. $|X'\Sigma^{-1}X|$ in (A.5) can be calculated as $|X'\Sigma^{-1}X| = |U'U|$.

If $\delta = 0$ and $\sigma_b^2 = 1$ in (A.4), we can apply the ordinary Kalman filter, given by the recursions

$$
\begin{aligned}
e_t &= z_t - H'\hat{x}_{t|t-1}, \qquad \sigma_{t|t-1}^2 = H'\Sigma_{t|t-1}H + \sigma_n^2 \\
K_t &= F\Sigma_{t|t-1}H/\sigma_{t|t-1}^2, \quad \hat{x}_{t+1|t} = F\hat{x}_{t|t-1} + K_t e_t \\
\Sigma_{t+1|t} &= (F - K_t H')\Sigma_{t|t-1}F' + GG',
\end{aligned}
$$

where the initial conditions are $\hat{x}_{1|0} = 0$ and $\Sigma_{1|0} = \Xi \mathrm{Var}(U)\Xi'$ and the covariance matrix $\mathrm{Var}(U)$ can be efficiently computed like in Jones (1980). The sequence $e_t/\sigma_{t|t-1}$, $t = 1, \ldots, N$ is an orthogonal sequence with mean zero and covariance matrix equal to the identity matrix. This implies that this sequence coincides with $L^{-1}z$ in (A.6). Therefore, the Kalman filter can be seen as an algorithm that, applied to any vector v of data, yields $L^{-1}v$. If δ is not zero in the GLS model (A.4), we can apply the Kalman filter to the data z and the columns of the X matrix to obtain $L^{-1}z$ and $L^{-1}X$. The DKF is an algorithm that computes these quantities automatically. In this algorithm, the recursions for e_t and $\hat{x}_{t|t-1}$ in the Kalman filter are augmented to matrix recursions

$$
\begin{aligned}
(e_t, E_t) &= (z_t, 0) - H'(\hat{x}_{t|t-1}, \hat{X}_{t|t-1}), \\
(\hat{x}_{t+1|t}, \hat{X}_{t+1|t}) &= F(\hat{x}_{t|t-1}, \hat{X}_{t|t-1}) + K_t(e_t, E_t),
\end{aligned}
$$

where the additional columns correspond to new states for the columns of the X matrix. The other recursions in the Kalman filter remain the same and the initialization is $(\hat{x}_{1|0}, \hat{X}_{1|0}) = (0, -A)$ and $\Sigma_{1|0}$ as before. It can be shown, using the results in De Jong (1991), that stacking the vectors (e_t, E_t) one on top of the other for $t = 1, \ldots, N$, the matrix $(L^{-1}z, L^{-1}X)$ is generated.

The DKF also has the recursion $Q_{t+1} = Q_t + (e_t, E_t)'(e_t, E_t)/\sigma_{t|t-1}^2$, initialized with $Q_1 = 0$. This recursion accumulates the partial squares and cross products in such a way that from Q_{N+1} the GLS estimators $\hat{\delta}$ and $\hat{\sigma}_b^2$ can computed. We propose a Kalman filter algorithm which is the DKF without the recursion for Q_t and which applies instead the QR algorithm to $(L^{-1}z, L^{-1}X)$, in the manner described above. We think that this procedure is numerically more stable than solving the normal equations to obtain the GLS estimators and is not computationally expensive.

Diffuse smoothing refers to the process of obtaining the estimator $\hat{x}_t$ of the state x_t based on the entire data vector $z = (z_1, \ldots, z_N)'$. The estimator $\hat{x}_t$ can be obtained by means of an augmented version of any of the existing algorithms for smoothing. We will use an augmented fixed point smoother, which for x_s, $1 \leq s \leq N$, is the set of recursions

$$K_t^a = \Sigma_{t|t-1}^a H/\sigma_{t|t-1}^2, \quad \Sigma_{t+1|t}^a = \Sigma_{t|t-1}^a(F - K_t H')'$$

$$(\hat{x}_{s|t}, \hat{X}_{s|t}) = (\hat{x}_{s|t-1}, \hat{X}_{s|t-1}) + K_t^a(e_t, E_t)$$
$$\Sigma_{s|t} = \Sigma_{s|t-1} - \Sigma_{t|t-1}^a H (K_t^a)',$$

initialized with $\Sigma_{s|s-1}^a = \Sigma_{s|s-1}$, where $\sigma_{t|t-1}^2$, K_t, (e_t, E_t), $(\hat{x}_{s|s-1}, \hat{X}_{s|s-1})$ and $\Sigma_{s|s-1}$ are produced by the proposed Kalman filter algorithm. It can be shown that the estimator $\hat{x}_s$ and its Mse are obtained from

$$\hat{x}_s = \hat{x}_{s|N} - \hat{X}_{s|N}\hat{\delta}, \quad \text{Mse}(\hat{x}_s) = \hat{\sigma}_b^2 \Sigma_{s|N} + \hat{X}_{s|N}\text{Mse}(\delta)\hat{X}_{s|N}'.$$

Details of Approach B

It is not difficult to verify that the Wiener–Kolmogorov formula corresponding to the signal s_t in (1) is given by

$$\hat{s}_t = \frac{\theta_s(B)\theta_s(F)\sigma_b^2}{\theta_s(B)\theta_s(F)\sigma_b^2 + \phi^*(B)\phi^*(F)\sigma_n^2} z_t. \tag{A.7}$$

Since the denominator in (A.7) is also $\theta(B)\theta(F)\sigma_a^2$, defining $k^2 = \sigma_b^2/\sigma_a^2$ and $\pi(B) = k\theta_s(B)/\theta(B)$, expression (A.7) can be written more compactly as $\hat{s}_t = \pi(B)\pi(F)z_t$.

The procedure used by Burman (1980) for signal extraction transforms the filter $\pi(B)\pi(F)$ into a sum of the form $\pi(B)\pi(F) = G(B) + G(F)$. This is the so–called parallel implementation of the filter. If the filter is applied to the series as a product of the two factors $\pi(B)$ and $\pi(F)$, this is called a cascade implementation. The cascade implementation is simpler than the parallel one since it is not necessary to partition the two–sided filter into two one–sided filters.

The algorithm for the cascade implementation can be obtained as follows. Let $y_t = \pi(B)z_t$. Then, using the ARIMA model for z_t and the definition of $\pi(B)$, it is easy to verify that y_t follows the model $\phi^*(B)y_t = k\theta_s(B)a_t$, where the a_t are the innovations of z_t. This, together with the fact that the series z_t also follows the backward model $\phi^*(F)z_t = \theta(F)v_t$, implies, after projecting onto the finite sample $z = (z_1, \ldots, z_N)'$, $\phi^*(B)y_t = 0$, $t \geq N + q_s + 1$, and $\phi^*(F)z_t = 0$, $t \leq -q$, where $q = max\{q_s, p+d\}$ is the degree of $\theta(B)$. Let p^* be the degree of $\phi^*(B)$. Then, the algorithm is

1. Solve the system

$$\begin{aligned} \theta(B)y_t &= k\theta_s(B)z_t & t &= -q+1, \ldots, p^* - q \\ \phi^*(F)y_t &= 0 & t &= -2q+1, \ldots, -q \end{aligned}$$

where $q + q_s$ backcasts are needed: $\hat{z}_{-q-q_s+1}, \ldots, \hat{z}_0$.
For $t = p^* - q + 1, \ldots, N + 2q_s$, obtain y_t from the recursion $\theta(B)y_t = k\theta_s(B)z_t$, where $2q_s$ forecasts are needed: $\hat{z}_{N+1}, \ldots, \hat{z}_{N+2q_s}$.

2. Solve the system

$$\begin{aligned} \theta(F)\hat{s}_t &= k\theta_s(F)y_t \quad t = N + q_s - p^* + 1, \ldots, N + q_s \\ \phi^*(B)\hat{s}_t &= 0 \qquad\qquad\;\; t = N + q_s + 1, \ldots, N + q_s + q \end{aligned}$$

For $t = N + q_s - p^*, \ldots, 1$, obtain $\hat{s}_t$ from the recursion $\theta(F)\hat{s}_t = k\theta_s(F)y_t$.

In order to obtain the forecasts and backcasts needed in step 1 of the previous algorithm, instead of using (A.1) and (A.2), it is easier to use a state space representation based on the reduced form ARIMA model $\phi^*(B)z_t = \theta(B)a_t$. The ordinary Kalman filter, initialized at $t = d + 1$, can be used like in Gómez and Maravall (1994) to compute the forecasts. Reversing the series and using the same procedure, the backcasts can also be obtained.

Details of Approach C

Suppose the observed series $z = (z_1, \ldots, z_N)'$ and let $s = (s_1, \ldots, s_N)'$. Without loss of generality, assume $\sigma_b^2 = 1$ and let $\lambda = \sigma_n^2/\sigma_b^2 = \sigma_n^2$. Let further $u_t = \alpha(B)s_t$, $u = (u_{d+1}, \ldots, u_N)'$ and $\mathrm{Var}(u) = \Omega$. Then, the problem is minimize $\sum_{t=1}^N (z_t - s_t)^2 + \lambda u'\Omega^{-1}u$. Define the $(N - d) \times N$ matrix

$$D = \begin{bmatrix} \alpha_d & \cdots & \alpha_1 & 1 & 0 & \cdots & 0 \\ 0 & \alpha_d & \cdots & \alpha_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \tag{A.8}$$

and let $\Omega = LL'$, with L lower triangular, be the Cholesky decomposition of Ω. Then, the problem can be expressed more compactly as minimize $n'n + \lambda(L^{-1}Ds)'L^{-1}Ds$, where $n = z - s$. Using standard matrix differentiation results, the solution can be easily seen to be

$$\hat{s}_t = [I + \lambda(L^{-1}D)'L^{-1}D]^{-1}z = \left[[I, \sqrt{\lambda}(L^{-1}D)'] \begin{bmatrix} I \\ \sqrt{\lambda}L^{-1}D \end{bmatrix}\right]^{-1} z. \tag{A.9}$$

In order to solve (A.9) efficiently, we can proceed as follows. Apply first the QR algorithm to the matrix $[I, \sqrt{\lambda}(L^{-1}D)']'$ to obtain an orthogonal $(2N-d)\times(2N-d)$ matrix Q such that $Q'[I, \sqrt{\lambda}(L^{-1}D)']' = [R', 0']'$, with R an upper triangular $N \times N$ matrix. Then, solve $R'R\hat{s}_t = z$.

The matrix $L^{-1}D$ can be computed bypassing the inversion of Ω or L by applying the Kalman filter corresponding to the model $u_t = \theta_s b_t$ to the columns of the matrix D. That is, after setting the state space representation corresponding to that model, the same Kalman filter is applied N times, using at iteration i the i-th column of D as data.

APPENDIX B: PROOFS

Proof of Theorem 2. It is not difficult to verify that $4\sin^2(x/2) = (1-e^{-ix})(1-e^{ix})$. Taking this expression into (5) and using the definition of λ, it is obtained that

$$|G(x)|^2 = \frac{1}{1+\lambda(1-e^{-ix})^d(1-e^{ix})^d}.$$

Replacing e^{-ix} with the backshift operator B in this expression implies that the filter is asserted.

In order to obtain the moving average polynomial $\theta(B)$ and σ_a^2 of the reduced form ARIMA, define $y = \sin^2(x/2)$. Then, $|G(x)|^2$ in (5) can be written as $|G(x)|^2 = C^d/(C^d + y^d)$ and the denominator in this expression can be factored as $(y - y_1)\cdots(y - y_d)$, where y_j can be expressed in polar form as $y_j = C_{(\pi+2j\pi)/2}$, $j = 0,\ldots,d-1$. If y_j is not real, then its conjugate $\bar{y}_j$ is also a root. Replacing e^{-ix} with B or, equivalently, y with $(1-B)(1-F)/4$, in $(y-y_j)(y-\bar{y}_j)$, an expression is obtained which can be factored as $\pi_j(B)\pi_j(F)/16$, where $\pi_j(B) = \alpha_{0,j} + \alpha_{1,j}B + \alpha_{2,j}B^2$. When d is even, all roots y_j are imaginary and, if d is odd, there are $d-1$ imaginary roots and one single real root $y_k = -C$. For this last root, $y + C$ can be factored as $\pi_k(B)\pi_k(F)/4$, where $\pi_k(B) = \alpha_{0,k} + \alpha_{1,k}B$. Therefore, the filter can be factored as $2^{2d}\sin(x_c/2)^{2d}/\prod_{j=1}^{k}(\pi_j(B)\pi_j(F))$, where k is the integer part of $(d+1)/2$. The expression for the $\alpha_{i,j}$, $i = 0,1,2$, corresponding to $(y-y_j)(y-\bar{y}_j)$, are obtained from the equalities

$$\alpha_{0,j}\alpha_{2,j} = 1, \qquad \alpha_{0,j}\alpha_{1,j}\alpha_{2,j} = -4 + 8C\cos((\pi + 2(j-1)\pi)/d)$$
$$\alpha_{0,j}^2 + \alpha_{1,j}^2 + \alpha_{2,j}^2 = 6 - 16C\cos((\pi + 2(j-1)\pi)/d) + 16C^2.$$

The expression for the $\alpha_{i,k}$, $i = 0,1$, corresponding to $y + C$ when d is odd, can be obtained analogously. □

Proof of Theorem 3. The proof is analogous to that of theorem 2, using the easily verifiable formula $\tan^2(x/2) = (1-e^{-ix})(1-e^{ix})/[(1+e^{-ix})(1+e^{ix})]$. □

References

[1] Akaike, H., (1974), "Markovian Representation of Stochastic Processes and its Applications to the Analysis of Autoregressive Moving Average Processes", *Annals of the Institute of Statistical Mathematics*, **26**, 363–387.

[2] Ansley, C.F. and Kohn, R., (1985) "Estimation, Filtering and Smoothing in State Space Models with Incompletely Specified Initial Conditions", *Annals of Statistics*, **13**, 1286–1316.

[3] Bell, W., (1984), "Signal Extraction for Nonstationary Series", *The Annals of Statistics*, **12**, 646-664.

[4] Box, G.E.P., and Jenkins, G.M., (1976), *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco.

[5] Brockwell, P., and Davis, R., (1992), *Time Series: Theory and Methods*, Springer-Verlag, Berlin.

[6] Burman, J.P., (1980), "Seasonal Adjustment by Signal Extraction", *Journal of the Royal Statistical Society* A, **143**, 321–337.

[7] Burns, A. M. and Mitchell, W. C., (1946), *Measuring Business Cycles*, New York, N.Y.: National Bureau of Economic Research, 1946.

[8] Dagum, E. B., (1980), "The X11 ARIMA Seasonal Adjustment Method", Statistics Canada, Catalogue 12–564E.

[9] De Jong, P., (1991), "The Diffuse Kalman Filter", *The Annals of Statistics*, **19**, 1073-1083.

[10] Gómez, V. (1997), "Three Equivalent Methods for Filtering Finite Nonstationary Time Series", submitted for publication.

[11] Gómez, V., and Maravall, A., (1994), "Estimation, Prediction and Interpolation for Nonstationary Series With the Kalman Filter", *Journal of the American Statistical Association*, **89**, 611-624.

[12] Gómez, V., and Maravall, A., (1996), "Programs TRAMO and SEATS", Instructions for the User (Beta Version: September 1996), Working Paper No. 9628, Bank of Spain.

[13] Hodrick, R. J. and Prescott, E. C. (1980), "Postwar U. S. business cycles: An empirical investigation", Discussion Paper No. 451, Carnegie–Mellon University.

[14] Jones, R., (1980), "Maximum Likelihood Fitting of ARMA Models to Time Series With Missing Observations", *Technometrics*, **22**, 389-395.

[15] King, R. G. and Rebelo, S. T. (1993), "Low Frequency Filtering and Real Business Cycles", *Journal of Economic Dynamics and Control*, **17**, 207–231.

[16] Kohn, R., and Ansley, C.F., (1985), "Efficient Estimation and Prediction in Time Series Regression Models", *Biometrika*, **72**, 694–697.

[17] Koopman, S. J., (1997), "Exact Initial Kalman Filter and Smoother for Nonstationary Time Series Models", Forthcoming in Journal of the American Statistical Association.

[18] Maravall, A., (1995), "Unobserved Components in Economic Time Series", in Pesaran, H., Schmidt, P. and Wickens, M. (eds.), *The Handbook of Applied Econometrics*, vol. 1, Oxford: Basil Blackwell.

[19] Oppenheim, A. V. and Schaffer, R. W., (1989), *Discrete–Time Signal Processing*, Prentice Hall, New Jersey.

[20] Otnes, R. K. and Enochson, L., (1978), *Applied Time Series Analysis*, Vol. 1, Wiley, New York.

[21] Whittaker, E. T. (1923), "On a new Method of Graduation", *Proc. Edinburgh Math. Assoc.*, **78**, 81–89.