# THREE EQUIVALENT METHODS FOR FILTERING FINITE NONSTATIONARY TIME SERIES

*Víctor Gómez*[*]

**D-98003**

**Febrero 1998**

(*) Ministerio de Economía y Hacienda

**Abstract**

To estimate the components in an unobserved ARIMA components model, three different approaches can be used: Kalman filtering plus smoothing, Wiener–Kolmogorov filtering and optimal smoothing. It is shown in the paper that the three approaches are equivalent. As an application, it is shown that any of the three approaches can be used to filter a series with the Hodrick–Prescott filter, since this filter can be given a signal extraction interpretation.

**Keywords:** Kalman filter; Signal extraction; ARIMA components model; Smoothing; Penalized Least Squares Smoothing; Wiener–Kolmogorov filters.

# 1 Introduction and Summary

In order to simplify the exposition, we will suppose first a time series $z = (z_1, \ldots, z_N)'$ that has been 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 $s_t$. The results obtained for model (1) will be extended later to the case in which $s_t$ can be decomposed in turn into several orthogonal components. We suppose that all components follow ARIMA models with mean zero and autoregressive polynomials, which include the roots of unit modulus, that are coprime. The series $z_t$ follows then the ARIMA model implied by the components, the so–called reduced form model.

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 algorithm is a simple modification of the diffuse Kalman filter of De Jong (1991), properly initialized.

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 $\alpha(B) = 1 + \alpha_1 B + \cdots + \alpha_d B^d$ be a polynomial in the the backshift operator, $B^k s_t = s_{t-k}$, with all its roots on the unit circle, which renders $s_t$ stationary. Let further $u_t = \alpha(B)s_t$, $t = d+1, \ldots, N$, $u = (u_{d+1}, \ldots, u_N)'$, $\mathrm{Var}(u) = \Omega$ and $\mathrm{Var}(n_t) = \lambda$ in (1). Then, the problem is minimize

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

It is to be noted that the only approach that allows for the computation of the mean squared errors (MSE) of the estimators is approach A. Using the

method of Gómez and Maravall (1996), 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), where the results of approach B cannot be applied.

Among the existing approaches to handle nonstationary state space models there are two that 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 and $z = (z_1, \ldots, z_N)'$ is the observed series, 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 is tantamount to replacing 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. However, to the best of this author's knowledge, in the finite nonstationary situation, the equivalence between Kalman filtering plus smoothing and Wiener–Kolmogorov filtering plus Tunnicliffe Wilson's algorithm, applied like in Burman (1980), remains an open question. In fact, the approach proposed by Burman (1980) lacked a sound theoretical foundation and the results of Bell (1984) were not applicable to finite nonstationary series. In this respect, Burridge and Wallis (1988) even stated that Wiener–Kolmogorov filtering could only be used with stationary series and that for nonstationary series one should use the Kalman filter.

According to Gersch and Kitagawa (1990), the use of approach C can be

traced back to Whittaker (1923). Whittaker 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. The properties of the solution are clear. If $\lambda = 0$, $s_t = z_t$ and the solution is a replica of the observations. As $\lambda$ becomes increasingly large, the smoothness constraint dominates the solution. Whittaker left the choice of $\lambda$ to the investigator.

One example of approach C is the Bayesian method used in the program BAYSEA, contained in the Fortran library TIMSAC–84, to decompose a time series into several components. The method is based on Akaike (1980a, 1980b). Another example is the filter proposed by Hodrick and Prescott (1980), hereafter referred to as HPF, 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 HPF can be given a signal extraction interpretation, whereby it 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 $H_{HP}(B, F)$

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

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

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

The question then immediately 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.

The paper shows that, under the appropriate assumptions, the three approaches yield the same result. Details of the algorithms used for the three approaches are given. In the case of approach A, a slight modification of

the DKF of De Jong (1991) is proposed which is numerically more stable. The initialization of the algorithm is obtained by a simple generalization to unobserved ARIMA components models of the initialization of Gómez and Maravall (1994a). For approach B, a cascade implementation is proposed instead of the parallel one used in Tunnicliffe Wilson's algorithm, which is easier to implement and produces the same results. To implement approach C, an efficient algorithm is proposed which uses the Kalman filter together with the $QR$ algorithm.

The structure of the paper is as follows. In Section 2, the three approaches are described in detail and their equivalence is established. Also, the results are extended to the case in which there are more than two components in (1). In Section 3, an example is given of the application of the techniques described in the paper to filtering economic time series with the Hodrick–Prescott filter.

# 2 Equivalence of the Three Approaches

In order to prove the equivalence of the three approaches, 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 $\alpha(B)$ has all its roots on the unit circle and degree $d$, the polynomial $\phi(B)$ has all its roots outside the unit circle and 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).

## 2.1 Details of Approach A

Among the state space representations of ARIMA models, we select that of Gómez and Maravall (1994a), 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

$$z_t = H'x_t + n_t \tag{4}$$

$$x_{t+1} = Fx_t + Gb_{t+1}, \tag{5}$$

where

$$F = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -\phi_r^* & -\phi_{r-1}^* & -\phi_{r-2}^* & \dots & -\phi_1^* \end{bmatrix}, \tag{6}$$

$x_t = (s_t, s_{t+1,t}, \dots, s_{t+r-1,t})'$, $H = (1, 0, \dots, 0)'$, $G = (1, \psi_1^*, \dots \psi_{r-1}^*)'$ and the $\psi_i^*$ weights are the coefficients obtained from $\psi^*(B) = \theta_s(B)/\phi^*(B) = \sum_{i=0}^{\infty} \psi_i^* B^i$. The elements of the state vector are defined as $s_{t+i,t} = s_{t+i} - \psi_0^* b_{t+i} - \cdots - \psi_{i-1}^* b_{t+1}$, $i = 1, \dots, r-1$. 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 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 $t > 0$, $1/\alpha(B) = \sum_{i=0}^{\infty} \xi_i B^i$ and the $A_t = (A_{1t}, \ldots, A_{dt})'$ can be recursively generated from

$$\begin{aligned} A_t &= (0, \ldots, 1, \ldots, 0), \qquad t = 1-d, \ldots, 0, \\ A_t &= -\alpha_1 A_{t-1} - \cdots - \alpha_d A_{t-d}, \quad t > 0, \end{aligned}$$

where for $t = 1-d, \ldots, 0$ the one is in the $(t+d)$–th position.

Like in Gómez and Maravall (1994a), p. 615, it can be shown that the initial state vector $x_1$ verifies $x_1 = A\delta + \Xi U$, where $A = [A_1, \ldots, A_r]'$, $\Xi$ 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 $x_1$, $\delta$ 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 $\delta$, which can be considered as a vector of nuisance random variables. Kalbfleisch and Sprott (1970) proposed several methods to eliminate the dependence of the likelihood on nuisance parameters, which are also valid in the present context. More specifically, the marginal likelihood, which is the likelihood of a transformation of the data to eliminate the nuisance parameters, is the approach proposed by Ansley and Kohn (1985). The Bayesian approach, which consists of considering $\delta$ diffuse, is the approach of De Jong (1991).

For algorithmical purposes, we will use the approach of De Jong (1991) in this paper. Using the transition equation (5), we have the following lemma, whose proof is straightforward and is omitted.

**Lemma 1** *Suppose that the series $z = (z_1, \ldots, z_N)'$ has been generated by the state space model (4) and (5), where $x_1 = A\delta + \Xi U$, as described earlier, and assume that $\delta$ is independent of the $n_t$ and the $b_t$. Then, the following representation holds*

$$z = X\delta + \epsilon, \tag{7}$$

*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 $\epsilon_t$, $t = 1, \ldots, N$, can be obtained from the recursions*

$$X_t' = H'J_t, \qquad J_{t+1} = FJ_t,$$

*with the initial condition* $J_1 = A$, *and*

$$\epsilon_t = n_t + H'\eta_t, \qquad \eta_{t+1} = F\eta_t + Gb_{t+1},$$

*with the initial condition* $\eta_1 = \Xi U$. *Besides,* $E(\epsilon) = 0$, *and* $Cov(\delta, \epsilon) = 0$.

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 (7). Following De Jong (1991), suppose that $\delta$ 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 $\delta$ and letting $l(z)$ be the log–likelihood of $z$ in (7) it is shown in De Jong (1991) that, apart from a constant, as $C^{-1} \to 0$,

$$\begin{aligned} l(z) + \frac{1}{2}\ln|\sigma_b^2 C| \to \; & - \frac{1}{2}\{(N-d)\ln(\sigma_b^2) + \ln|\Sigma| + \ln|X'\Sigma^{-1}X| \\ & + (z - X\hat{\delta})'\Sigma^{-1}(z - X\hat{\delta})/\sigma_b^2\}, \end{aligned} \tag{8}$$

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 (8) is the diffuse log–likelihood. The parameter $\sigma_b^2$ can be concentrated out of the diffuse log–likelihood by replacing $\sigma_b^2$ in (8) with its maximum likelihood estimator $\hat{\sigma}_b^2$ $= (z - X\hat{\delta})'\Sigma^{-1}(z - X\hat{\delta})/(N-d)$.

The previous result tells us that making $\delta$ diffuse implies that (7) can be considered as a generalized linear regression model (GLS), where $\delta$ is the vector of regression parameters and $\hat{\delta}$ and $\hat{\sigma}_b^2$ are the GLS estimators.

One interesting point to note is that the diffuse log–likelihood (8) coincides with the log–likelihood of Box and Jenkins (1976) corresponding to the reduced form ARIMA model of $\{z_t\}$. This can be seen by considering that, as shown in Gómez and Maravall (1994b), p.49, the diffuse likelihood and the marginal likelihood of Ansley and Kohn (1985) coincide, and that the marginal likelihood is the likelihood of the differenced series (the Box and Jenkins' likelihood) or a generalization of it.

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 (7) by $L^{-1}$, it is obtained that

$$L^{-1}z = L^{-1}X\delta + L^{-1}\epsilon, \tag{9}$$

where $\text{Var}(L^{-1}\epsilon) = \sigma_b^2 I_N$. Therefore, model (9) 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 $(R', 0')' = QL^{-1}X$, where $R$ is a nonsingular $d \times d$ upper triangular matrix. Then, $\hat{\delta} = R^{-1}v_1$ 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 (8) can be calculated as $|X'\Sigma^{-1}X| = |R'R|$.

To understand the meaning of the DKF of De Jong (1991), consider first that $\delta$ is zero in (7) and $\sigma_b^2 = 1$. Then, we can apply the ordinary Kalman filter, given by the recursions

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

where the initial conditions are $\hat{x}_{1|0} = 0$ and $\Sigma_{1|0} = \Xi\text{Var}(U)\Xi'$ and the covariance matrix $\text{Var}(U)$ can be efficiently computed like in Jones (1980). The sequence of standardized innovations $e_t/\sigma_{t|t-1}$, $t = 1, \ldots, N$ 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 (9). Also, $|L| = \prod_{t=1}^{N} \sigma_{t|t-1}$. These are standard results of the Kalman filter. Proofs can be seen in Anderson and Moore (1979).

A consequence of these results is that the Kalman filter can be seen as an algorithm that, applied to any vector $v$ of data, yields $L^{-1}v$. Therefore, if $\delta$ is not zero in the GLS model (7), 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 allows for the automatic computation of these quantities. 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 $(e_t, E_t)/\sigma_{t|t-1}$ 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 $Q_{t+1} = Q_t + (e_t, E_t)'(e_t, E_t)/\sigma^2_{t|t-1}$, initialized with $Q_1 = 0$. This recursion accumulates the partial squares and cross products in such a way that

$$Q_{N+1} = \begin{bmatrix} (L^{-1}z)' \\ (L^{-1}X)' \end{bmatrix} \left[L^{-1}z, L^{-1}X\right] = \begin{bmatrix} z'\Sigma^{-1}z & z'\Sigma^{-1}X \\ X'\Sigma^{-1}z & X'\Sigma^{-1}X \end{bmatrix}$$

and from $Q_{N+1}$ the GLS estimators $\hat{\delta}$ and $\hat{\sigma}^2_b$ can be computed. We propose in this paper 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.

Note that $\sigma^2_b$ is supposed to be one in the proposed algorithm because it can be estimated later with GLS. Instead of concentrating out of the diffuse likelihood this parameter, we could have concentrated out $\sigma^2_n$. In this case, we would use $\sigma^2_b/\sigma^2_n$ instead of $\sigma^2_b$ in the proposed procedure.

Once the GLS estimators $\hat{\delta}$ and $\hat{\sigma}^2_b$ in (7) have been obtained, it can be shown, using the results in De Jong (1991), that the diffuse predictors $\hat{z}_{N+1}$ and $\hat{x}_{N+1}$ of $z_{N+1}$ and $X_{N+1}$ are

$$\hat{z}_{N+1} = H'\hat{x}_{N+1|N} + E_{N+1}\hat{\delta}, \quad \hat{x}_{N+1} = \hat{x}_{N+1|N} - \hat{X}_{N+1|N}\hat{\delta}$$
$$\mathrm{Mse}(\hat{z}_{N+1}) = \hat{\sigma}^2_b\sigma^2_{t|t-1} + E_{N+1}\mathrm{Mse}(\hat{\delta})E'_{N+1}$$
$$\mathrm{Mse}(\hat{x}_{N+1}) = \hat{\sigma}^2_b\Sigma_{N+1|N} + \hat{X}_{N+1|N}\mathrm{Mse}(\hat{\delta})\hat{X}'_{N+1|N},$$

where $\mathrm{Mse}(\hat{\delta}) = \hat{\sigma}^2_b(X'\Sigma^{-1}X)^{-1} = \hat{\sigma}^2_b R^{-1}R'^{-1}$.

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, like the fixed point smoother or the fixed interval smoother. In this paper we will use an augmented fixed point smoother because it can be simplified so that very small storage requirements are needed, see Gómez and Maravall (1994a), and because it is well suited for revisions of the estimates as new data come in.

The augmented fixed point smoother 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

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

We emphasize that we do not claim that the proposed algorithms for Kalman filtering and smoothing are the best ones available. However, they are efficient and numerically stable, without being computationally expensive.

EXAMPLE 1 Suppose a yearly univariate series $z = (z_1, \ldots, z_N)'$ which follows the model $z_t = s_t + n_t$, where the model for the signal $s_t$ is $\nabla s_t = b_t$. Define the state vector $x_t = s_t$. Then, a state space representation is given by (4) and (5), where $F = 1$, $G = 1$, $H = 1$, and $\delta = s_0$. The initial state is $x_1 = \delta + b_1$ and $\sigma_b^2$ can be concentrated out of the likelihood by defining $\text{Var}(b_t', n_t')' = \sigma_b^2 \text{ diag}(1, \lambda)$, where $\lambda = \sigma_n^2/\sigma_b^2$. The proposed algorithm is initialized with $(\hat{x}_{1,0}, \hat{X}_{1,0}) = (0, -1)$ and $\Sigma_{1,0} = 1$.

EXAMPLE 2 The Hodrick–Prescott filter. Suppose a quarterly univariate series $z = (z_1, \ldots, z_N)'$ which follows the model $z_t = s_t + n_t$, where the model for the signal $s_t$ is $\nabla^2 s_t = b_t$. The state vector is defined as $x_t = (s_t, s_{t|t-1})'$, the polynomial $\alpha(B) = \nabla^2$, where $d = 2$, and the degrees of $\phi(B)$ and $\theta_s(B)$ are $p = q_s = 0$. Then, $r = 2$, $\phi^*(B) = \alpha(B)$, and the state space representation is given by by (4) and (5), where $\phi_2^* = -1$ and $\phi_1^* = 2$ in (6), $G = (1, 2)$, $H = (1, 0)$ and $\delta = (s_{-1}, s_0)'$. The initial state is

$$x_1 = \begin{bmatrix} -1 & 2 \\ -2 & 3 \end{bmatrix} \delta + \begin{bmatrix} 1 \\ 2 \end{bmatrix} b_1.$$

$\sigma_b^2$ can be concentrated out of the likelihood by defining $\text{Var}(b_t', n_t')' = \sigma_b^2$ $\text{diag}(1, \lambda)$, where $\lambda = \sigma_n^2/\sigma_b^2$. Therefore, we initialization of the proposed

algorithm is

$$(\hat{x}_{1|0}, \hat{X}_{1|0}) = \begin{bmatrix} 0 & 1 & -2 \\ 0 & 2 & -3 \end{bmatrix}, \qquad \Sigma_{1|0} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}.$$

## 2.2 Details of Approach B

Under the assumptions and with the notation of this Section, it is not difficult to verify that the Wiener–Kolmogorov formula corresponding to the signal $s_t$ in (1) is given by

$$\hat{s}_t = \frac{\theta_s(B)\theta_s(F)\sigma_b^2}{\theta_s(B)\theta_s(F)\sigma_b^2 + \phi^*(B)\phi^*(F)\sigma_n^2} z_t. \tag{10}$$

Since the denominator in (10) 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 (10) 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 $\pi(B)\pi(F) = G(B) + G(F)$. Using the filter jargon, this can be described as a 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. That is, the filtered series $\hat{s}_t$ is obtained as $\hat{s}_t = \pi(F)\,[\pi(B)z_t]$. 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 $z = (z_1, \ldots, z_N)'$,

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

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

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

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 (4) and (5), 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 (1994a) to compute the forecasts. Reversing the series and using the same procedure, the backcasts can also be obtained.

## 2.3 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

$$D = \begin{bmatrix} \alpha_d & \cdots & \alpha_1 & 1 & 0 & \cdots & 0 \\ 0 & \alpha_d & \cdots & \alpha_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \tag{11}$$

and let $\Omega = LL'$, with $L$ lower triangular, be the Cholesky decomposition of $\Omega$. 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

$$\hat{s}_t = [I + \lambda(L^{-1}D)'L^{-1}D]^{-1}z = \left[[I, \sqrt{\lambda}(L^{-1}D)'] \begin{bmatrix} I \\ \sqrt{\lambda}L^{-1}D \end{bmatrix}\right]^{-1} z. \quad (12)$$

In order to solve (12) in a numerically safe manner, 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 $\Omega$ 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.

Note that the matrices involved in this approach may be of considerable dimension and, therefore, it is important to use a computational procedure that is numerically stable.

## 2.4 Equivalence of Approaches A, B and C

The main result of this Section is contained in the following theorem. The proof is in Appendix B.

**Theorem 1** *Suppose that the vector $\delta$ and the proposed Kalman filter algorithm and augmented fixed point smoother are used when approach A is applied. Then, under the assumptions of this Section, the approaches A, B and C are equivalent.*

It is to be noted that, as far as the estimator of $s_t$ based on the finite sample is concerned, it is irrelevant which of the two assumptions of Bell (1984), A or B, is made. This follows from result 1 of Bell and Hillmer (1991) and the transformation of Bell (1984), p. 651, between the starting values of both assumptions.

The methods described in this Section to implement approaches A, B and C will be applied to an example in Appendix C, where all details of the computations will be given.

## 2.5 Extension of Results

The results of this Section can be extended without difficulty to more than two components. Consider, for example, the decomposition $z_t = p_t + s_t + w_t$, where $p_t$ is the trend, $s_t$ is the seasonal and $w_t$ is the irregular component. Then, approach A can be implemented using the diffuse likelihood approach and the proposed Kalman filter algorithm of Section 2.1. Only a simple generalization of both the state space representation and the initialization to two components is needed. Specifically, the state vector is defined as $x_t = (x_t^{p'}, x_t^{s'})'$, where $x_t^p$ and $x_t^s$ correspond to $p_t$ and $s_t$ and are defined in terms of the models followed by these components, in a manner similar to that used in Section 2.1 for $s_t$. Using the two subcomponents $x_t^p$ and $x_t^s$, the initialization is now obvious.

Approach B was already applied in Burman (1980) to more than two components and approach C can be applied using the results of Akaike (1980a, 1980b). More specifically, let $\alpha_p(B)$ and $\alpha_s(B)$ be polynomials with all their roots on the unit circle of degrees $m$ and $n$, $u_t^p = \alpha_p(B)p_t$, $u^p = (u_{m+1}^p, \ldots, u_N^p)'$, $u_t^s = \alpha_s(B)s_t$, $u^s = (u_{n+1}^s, \ldots, u_N^s)'$, $w_t \sim N(0, \sigma^2)$, $u^p \sim N(0, \sigma^2\lambda_p\Omega_p)$, and $u^s \sim N(0, \sigma^2\lambda_s\Omega_s)$. Then, the problem is minimize

$$\sum_{t=1}^{N}(z_t - p_t - s_t)^2 + \frac{1}{\lambda_p}u^{p'}\Omega_p^{-1}u^p + \frac{1}{\lambda_s}u^{s'}\Omega_s^{-1}u^s.$$

The normality assumption has been made to simplify the exposition, but the results are also true for the nonnormality case. The only difference would be that the estimators would be the best ones in the mean squared sense only.

# 3 Application

Several programs in Fortran have been written by the author to implement the methodology outlined in the paper. The programs allow for the application of the three approaches and 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.

As an example, we have used 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 and the parameter estimates to SEATS, signal extraction is performed. Then, the HPF is applied to the seasonally adjusted series (SAS) using approaches A, B and C. The results, only for the first year and the last year of the data, are shown in table 1. They are practically identical for the three approaches and the same thing happens with the rest of the data.

The gain functions of the HPF and the trend and SAS filters used by SEATS are displayed in figure 1(a), whereas, in figure 1(b), one can see the trend component estimated by SEATS and the smoother trend obtained by filtering the SAS with the HPF. In the x–axis in figure 1(a) we have frequencies, from zero to $\pi$. In figure 1(b), in the x–axis we have observations and in the y–axis we have the trends in logarithms. The HPF 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).

Table 1: US GNP filtered with the Hodrick–Prescott filter

| | | Approach A | Approach B | Approach C |
|---|---|---|---|---|
| 1951 | I | 4.397693554171079 | 4.397693554171953 | 4.397693554170885 |
| | II | 4.409045168767950 | 4.409045168768817 | 4.409045168767809 |
| | III | 4.420389562455971 | 4.420389562456831 | 4.420389562455878 |
| | IV | 4.431715093283309 | 4.431715093284160 | 4.431715093283252 |
| 1985 | I | 6.879890196137601 | 6.879890196139041 | 6.879890196137391 |
| | II | 6.900816816259312 | 6.900816816260792 | 6.900816816259041 |
| | III | 6.921724398682671 | 6.921724398684195 | 6.921724398682327 |
| | IV | 6.942624409350197 | 6.942624409351771 | 6.942624409349766 |

## APPENDIX A

**Proof of Theorem 1.**

We prove first the equivalence of approaches A and B. Suppose that we want to use approach A to estimate $s_t$ based on $z = (z_1, \ldots, z_N)'$, the observed series. To this end, the proposed Kalman filter is first applied, followed by the $QR$ algorithm, to obtain the GLS estimator $\hat{\delta}$. Then, we apply the augmented fixed point smoother. Denote this estimator by $\mathrm{E}(s_t|z, \hat{\delta})$.

Suppose now that $\eta = (z_{1-d}, \ldots, z_0)'$ is used instead of $\delta$ to model uncertainty and that approach A is used again to estimate $s_t$ based on $z$. Denote this estimator by $\mathrm{E}(s_t|z, \hat{\eta})$. Then, by result 1 and examples 1 and 2 of Bell and Hillmer (1991), the transformation approach estimates of $s_t$ using $\delta$ and $\eta$ coincide. By theorem 5.2 of Ansley and Kohn (1985), these last two estimators also coincide with the diffuse estimators $\mathrm{E}(s_t|z, \hat{\delta})$ and $\mathrm{E}(s_t|z, \hat{\eta})$.

Make assumption A of Bell (1984) and suppose that $\eta$ instead of $z_* = (z_1, \ldots, z_d)'$ is used to generate the series, which in the present context means that $\eta$ is independent of $\{b_t\}$ and $\{n_t\}$, and assume that the complete realization $\{\ldots, z_{-1}, z_0, z_1, \ldots\}$ is known. Then, by the results of Bell (1984), pp. 662–663, the Wiener–Kolmogorov filter can be applied to the doubly–infinite series to obtain the best linear estimator, in the mean squared sense, of $s_t$, which will be denoted by $\mathrm{E}(s_t|\{z_t\})$. Projecting first this estimator onto the space generated by $\{\eta, z_1, \ldots, z_N\}$, where $\eta$ is considered fixed, and then projecting this projector onto the space generated by $\{z_1, \ldots, z_N\}$, yields $\mathrm{E}(s_t|z, \hat{\eta})$, the estimator mentioned above. This implies that $\mathrm{E}(s_t|z, \hat{\delta})$ coincides with the estimator that results after replacing in $\mathrm{E}(s_t|\{z_t\})$ the unknown $z_t$ with forecasts and backcasts, which can be obtained using the reduced form ARIMA model and an appropriate state

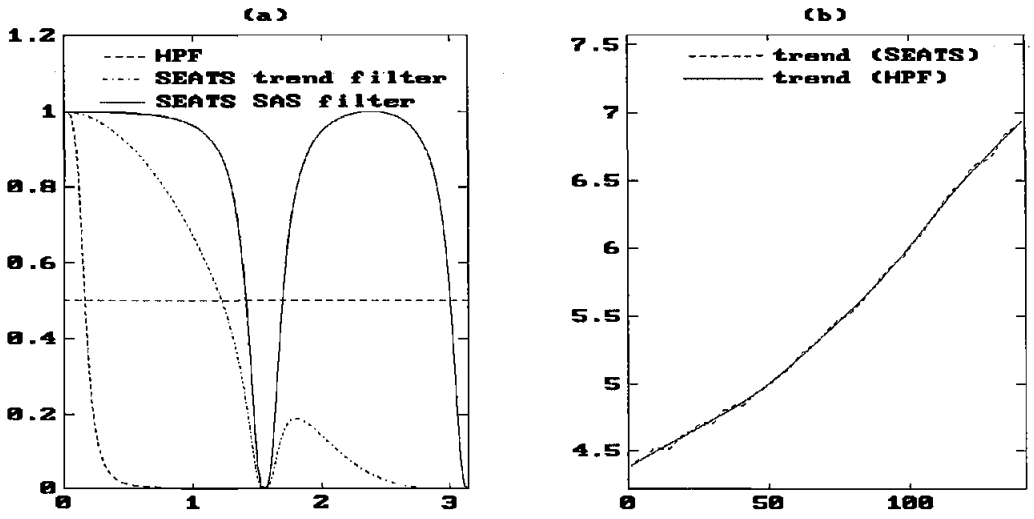


Figure 1: (a) Gain Functions of HPF and SEATS Trend and SAS Filters. (b) Trends estimated by SEATS and HPF.

space representation. The algorithm of G. Tunnicliffe Wilson can be used like in Burman (1980) to avoid the need to compute a large number of forecasts and backcasts.

To prove the equivalence of approaches A and C, let $s = (s_1, \ldots, s_N)'$ and partition $s = (s_I', s_{II}')'$, where $s_I = (s_1, \ldots, s_d)'$ and $s_{II} = (s_{d+1}, \ldots, s_N)'$. Choose $\gamma = s_I$ to model uncertainty and let further $u_t = \alpha(B)s_t$, $t = d+1, \ldots, N$, $u = (u_{d+1}, \ldots, u_N)'$, and $n = (n_1, \ldots, n_N)'$. Assume that $\gamma$ is independent of $u$ and $n$ and that $s$ and $n$ are normally distributed, and consider the joint distribution of $(s, z, \gamma)$. Since $s_I = \gamma$, the density $p(s, z, \gamma)$ is a degenerate density and exists in the subspace generated by the variables contained in $(s, z)$. Therefore, $p(s, z, \gamma) = p(s, z)$ and we can write the following equality between densities

$$p(s|z, \gamma)p(z|\gamma)p(\gamma) = p(z|s)p(s),$$

where $p(s|z, \gamma)$ is also a degenerate density which coincides with $p(s_{II}|z, \gamma)$. On the other hand, we have $p(s) = p(s_{II}|s_I)p(\gamma)$ and, by lemma 1 of Gómez and Maravall (1994a), $p(s_{II}|s_I) = p(u)$. Substituting into the previous expression and cancelling terms yields

$$p(s|z, \gamma)p(z|\gamma) = p(z|s)p(u).$$

The equality between densities implies

$$(s - \mathrm{E}(s|z,\gamma))'\Omega^{-}_{s|z,\gamma}(s - \mathrm{E}(s|z,\gamma)) + (z - X\gamma)'\Omega^{-1}_{z|\gamma}(z - X\gamma)$$
$$= (z - s)'(z - s)/\lambda + u'\Omega_u^{-1}u,$$

where $\Omega_{s|z,\gamma}$, $\Omega_{z|\gamma}$, and $\Omega_u$ are the covariance matrices of $p(s|z,\gamma)$, $p(z|\gamma)$ and $p(u)$. Writing $u = Ds$, where $D$ is the matrix defined in (11), and premultiplying the previous equality by $\lambda$, it is obtained that

$$\lambda\left[(s - \mathrm{E}(s|z,\gamma))'\Omega^{-}_{s|z,\gamma}(s - \mathrm{E}(s|z,\gamma)) + (z - X\gamma)'\Omega^{-1}_{z|\gamma}(z - X\gamma)\right]$$
$$= (z - s)'(z - s) + \lambda s'D'\Omega_u^{-1}Ds.$$

The value of $s$ which minimizes the left hand side of the previous equality must be equal to the one that minimizes the right hand side. A brief inspection of the left hand side shows that the second term is minimized when $\gamma$ is equal to $\hat{\gamma}$, the GLS estimator of model (7). For any $\gamma$, the first term of the left hand side is minimized for $s$ equal to $\mathrm{E}(s|z,\gamma)$. Therefore, the minimizer of the left hand side is the result of replacing $\gamma$ in $\mathrm{E}(s|z,\gamma)$ with $\hat{\gamma}$. It is now evident that this minimizer coincides with $\mathrm{E}(s|z,\hat{\gamma})$, the estimator of $s$, based on $z$ and $\gamma$, obtained with approach A. By result 1 and example 2 of Bell and Hillmer (1991), $\mathrm{E}(s|z,\hat{\gamma})$ coincides with $\mathrm{E}(s|z,\hat{\delta})$, the estimator obtained with approach A, based on $z$ and $\delta$. Using calculus, it is easy to see that the minimizer of the right hand side is $\hat{s} = (I + \lambda D'\Omega_u^{-1}D)^{-1}z$. □

## APPENDIX B

In this appendix, approaches A, B and C will be applied to example 1, with the particular values $\sigma_n^2 = 2$, $\sigma_b^2 = 1$, and supposing the observed series is $z = (z_1, z_2, z_3)'$. Details of the computations will be given.

Since $z_t = s_t + n_t$ and $s_t$ follows the model $\nabla s_t = b_t$, the series $z_t$ follows the model $\nabla z_t = (1 + \theta B)a_t$ and the parameters $\theta$ and $\sigma_a^2 = \mathrm{Var}(a_t)$ can be obtained by means of the autocovariance generating function corresponding to $(1 + \theta B)a_t = b_t + (1 - B)n_t$. Performing the calculations yields $\sigma_a^2 = 4$ and $\theta = -1/2$.

Suppose approach A is applied first and define the state vector $x_t = s_t$. Then, as shown in Section 2, a state space representation is given by (4) and (5), where $F = 1$, $G = 1$, $H = 1$, and $\delta = s_0$. The initial state is $x_1 = \delta + b_1$ and the proposed Kalman filter algorithm is initialized with $(\hat{x}_{1|0}, \hat{X}_{1|0}) = (0, -1)$ and $\Sigma_{1|0} = 1$.

Applying the proposed Kalman filter for $t = 1, 2, 3$, yields $(e_1, E_1) = (z_1, 1)$, $\sigma^2_{1|0} = 3$, $(\hat{x}_{2|1}, \hat{X}_{2|1}) = (z_1/3, -2/3)$, $\Sigma_{2|1} = 5/3$, $(e_2, E_2) = (-z_1/3 + z_2, 2/3)$,

$\sigma^2_{2|1} = 11/3$, $(\hat{x}_{3|2}, \hat{X}_{3|2}) = (2z_1/11 + 5z_2/11, -4/11)$, $\Sigma_{3|2} = 21/11$, $(e_3, E_3) = (-2z_1/11 - 5z_2/11 + z_3, 4/11)$, $\sigma^2_{3|2} = 43/11$.

Using the notation of section 2.1, the matrices $L^{-1}z$ and $L^{-1}X$ of model (9) are $L^{-1}z = (z_1/\sqrt{3}, (-z_1/3 + z_2)\sqrt{3/11}, (-2z_1/11 - 5z_2/11 + z_3)\sqrt{11/43})'$ and $L^{-1}X = (1/\sqrt{3}, (2/3)\sqrt{3/11}, (4/11)\sqrt{11/43})'$. From this, it is obtained that $\hat{\delta} = 11z_1/21 + 6z_2/21 + 4z_3/21$ and $\text{Mse}(\hat{\delta}) = 43/21$.

Applying the fixed point smoother to estimate $x_1$ yields $\Sigma^a_{1|0} = 1$, $\Sigma^a_{2|1} = 2/3$, $(\hat{x}_{1|1}, \hat{X}_{1|1}) = (z_1/3, -2/3)$, $\Sigma^a_{3|2} = 4/11$, $(\hat{x}_{1|2}, \hat{X}_{1|2}) = (3z_1/11 + 2z_2/11, -6/11)$, $(\hat{x}_{1|3}, \hat{X}_{1|3}) = (11z_1/43 + 6z_2/43 + 4z_3/43, -22/43)$. Using the estimator $\hat{\delta}$, it is obtained that $\hat{s}_1 = \hat{x}_{1|3} - \hat{X}_{1|3}\hat{\delta} = 11z_1/21 + 2z_2/7 + 4z_3/21$. Proceeding similarly, the fixed point smoother yields $\hat{s}_2 = 2z_1/7 + 3z_2/7 + 2z_3/7$ and $\hat{s}_3 = 4z_1/21 + 2z_2/7 + 11z_3/21$.

In order to apply approach B, let $k = 1/\sigma_a = 1/2$. Then, $\hat{s}_t = H(F)H(B)z_t$, where $H(B) = k/\theta(B)$. Using the cascade implementation, the following two systems have to be solved

$$(1 - B/2)y_t = z_t/2 \quad t = 0 \tag{B.1}$$

$$(1 - F)y_t = 0 \quad t = -1 \tag{B.2}$$

and

$$(1 - F/2)\hat{s}_t = y_t/2 \quad t = 3 \tag{B.3}$$

$$(1 - B)\hat{s}_t = 0 \quad t = 4. \tag{B.4}$$

One backcast $\hat{z}_0$ is necessary for the previous systems. This can be obtained by first reversing the series and then applying the Kalman filter like in Gómez and Maravall (1994a). A suitable state space representation is given by $z_t = H'_z x_t + a_t$ and $x_{t+1} = F_z x_t + G_z a_t$, where $x_t = z_{t,t-1} = z_t - a_t$, $H'_z = 1$, $F_z = 1$ and $G_z = 1 + \theta$. Using the reversed series $(z_3, z_2, z_1)'$, choose $\eta = z_3$ to model uncertainty with respect to the initial conditions and take $x_2 = n\eta + \theta a_1$ as first state. Then, the Kalman filter can be used, initialized with $\hat{x}_{2,1} = z_3$ and $\Sigma_{2,1} = \theta^2$, where $\sigma^2_a$ is supposed to be one because it is concentrated out of the likelihood. Performing the calculations yields $\hat{z}_0 = 11z_1/21 + 2z_2/7 + 4z_3/21$.

Solving first (B.1) and (B.2), it is obtained that $y_{-1} = y_0 = \hat{z}_0$. Using then the recursion $y_t = y_{t-1}/2 + z_t/2$ for $t = 1, 2, 3$, yields $y_1 = 16z_1/21 + z_2/7 + 2z_3/21$, $y_2 = 8z_1/21 + 4z_2/7 + z_3/21$ and $y_3 = 4z_1/21 + 2z_2/7 + 11z_3/21$. The solution of (B.3) and (B.4) is $\hat{s}_3 = \hat{s}_4 = y_3$. Using the recursion $\hat{s}_t = \hat{s}_{t+1}/2 + y_t$ for $t = 2, 1$, it is seen that the results coincide with those of approach A.

Suppose approach C is applied. Then, the problem is minimize $\sum_{t=1}^{3}(z_t - s_t)^2 + 2\sum_{t=2}^{3}(\nabla s_t)^2$. Letting $s = (s_1, s_2, s_3)'$, the solution is given by $\hat{s} = (I + 2D'D)^{-1}z$,

where $D$ is the $2 \times 3$ matrix

$$D = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}.$$

It is easy to check that

$$(I + 2D'D)^{-1} = \begin{bmatrix} 11/21 & 6/21 & 4/21 \\ 2/7 & 3/7 & 2/7 \\ 4/21 & 2/7 & 11/21 \end{bmatrix}.$$

# 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] Akaike, H., (1980a), "Likelihood and the Bayes Procedure ", in *Bayesian Statistics*, eds., J.M. Bernardo, M.H. De Groot, D.V. Lindley and A.F.M. Smith, University Press, Valencia, Spain, 143–166.

[3] Akaike, H., (1980b), "Seasonal Adjustment by a Bayesain Modeling", *Journal of Time Series Analysis*, **1**, 1–13.

[4] Anderson, B.D.O., and Moore, J.B., (1979), *Optimal Filtering*, Englewood Cliffs: Prentice Hall.

[5] 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.

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

[7] Bell, W., and Hillmer, S., (1991), "Initializing the Kalman Filter for Nonstationary Time Series Models", *Journal of Time Series Analysis*, **12**, 283–300.

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

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

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

[11] Burridge, P. and Wallis K. F., (1988), "Prediction Theory for Autoregressive Moving Average Processes", *Econometric Reviews*, **7(1)**, 65–95.

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

[13] Gersch, W. and Kitagawa, G, (1988), "Smoothness Priors in Time Series", in J. C. Spall (ed.), *Bayesian Analysis of Time Series and Dynamic Models*, Marcel Dekker, Inc., New York.

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

[15] Gómez, V., and Maravall, A., (1994b), "Initializing the Kalman Filter with Incompletely Specified Initial Conditions", in Chen, G.R. (ed.), *Approximate Kalman Filtering (Series on Approximation and Decomposition)*, London: World Scientific Publ. Co.

[16] 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.

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

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

[19] Kalbfleisch, J. D., and Sprott, D. A., (1970), "Application of Likelihood Methods to Models Involving Large Numbers of Parameters" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, **32**, 175-208.

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

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

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

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