Biometrika (1993), **80**, 1, *pp.* 117-26
Printed in Great Britain

Disturbance smoother for state space models

BY SIEM JAN KOOPMAN
Department of Statistical and Mathematical Sciences,
London School of Economics and Political Science,
Houghton Street, London WC2A 2AE, U.K.

SUMMARY

This paper develops a method to evaluate the smoothed estimator of the disturbance vector in a state space model together with its mean squared error matrix. This disturbance smoother also leads to an efficient smoother for the state vector. Applications include a method to calculate auxiliary residuals for unobserved components time series models and an EM algorithm for estimating covariance parameters in a state space model.

Some key words: Disturbance smoother; EM algorithm; Kalman filter; Residual; State smoother; State space model; Unobserved components time series model.

1. INTRODUCTION

Assume that a vector of time series y_t is generated by the state space model

$$y_t = Z_t\alpha_t + X_t\beta + G_t\varepsilon_t, \quad \beta = b + B\delta \quad (t = 1, \ldots, n), \tag{1·1}$$

$$\alpha_{t+1} = T_t\alpha_t + W_t\beta + H_t\varepsilon_t, \quad \alpha_0 = 0 \quad (t = 0, \ldots, n), \tag{1·2}$$

where the vector b and the matrix B are known and the unknown parameter vector δ can be regarded as fixed or diffuse (de Jong, 1991). Under Gaussianity, the disturbance vectors ε_t $(t = 0, \ldots, n)$ are a sequence of independent normal random variables with mean zero and covariance matrix $\sigma^2 I$, that is $\varepsilon_t \sim \text{NID}(0, \sigma^2 I)$. The first equation of the state space model is referred to as the measurement equation whereas the second is called the transition equation. The system matrices are Z_t, X_t, G_t, T_t, W_t and H_t.

A variety of time series models can be cast in the state space model (1·1)-(1·2): regression effects in time series models are placed in $X_t\beta$; autoregressive moving average dynamics are found in $T_t\alpha_t$ and $H_t\varepsilon_t$; additive outlier and structural change interventions are modelled by using $X_t\beta$ and $W_t\beta$, respectively; unobserved components are placed in the state vector α_t; and the covariance structure is determined by the matrices G_t and H_t. The state space model can be regarded as a natural set up for a wide range of time series models in order to estimate parameters and to predict future observations.

We set up the Kalman filter to produce, at time $t-1$, the estimator of the state vector at time t where the parameter vector δ is assumed to be known so that the initial state vector and the regression effects are also known. For a Gaussian state space model, the minimum mean squared error estimator of the state vector, based on $Y_{t-1} = \{y_1, \ldots, y_{t-1}\}$ and parameter δ, is $a_t = E(\alpha_t | Y_{t-1}, \delta)$. When the Gaussian assumption does not hold, vector a_t is the minimum mean squared error linear estimator. The mean squared error matrix of a_t is written as $\sigma^2 P_t = \text{MSE}(a_t)$. It follows that the one-step ahead prediction error $v_t = y_t - E(y_t | Y_{t-1}, \delta)$ is given by

$$v_t = y_t - Z_t a_t - X_t\beta, \quad F_t = Z_t P_t Z_t' + G_t G_t' \quad (t = 1, \ldots, n),$$

where $\sigma^2 F_t$ is the covariance matrix of v_t. The updating equations for a_t and P_t are

$$a_{t+1} = T_t a_t + W_t \beta + K_t v_t, \quad P_{t+1} = T_t P_t L_t' + H_t M_t' \quad (t = 1, \ldots, n),$$

where

$$K_t = (T_t P_t Z_t' + H_t G_t') F_t^{-1}, \quad L_t = T_t - K_t Z_t, \quad M_t = H_t - K_t G_t.$$

The recursions are started off with $a_1 = W_0 \beta$ and $P_1 = H_0 H_0'$. A straightforward proof of the Kalman filter is given by Anderson & Moore (1979, p. 105). To allow for the unknown vector δ, de Jong (1991) proposes to augment the Kalman filter equations for v_t and a_t; see Appendix 1.

While the Kalman filter uses past observations, a state smoother yields estimators of the state vector based on the full set of observations in the sample. The classical fixed interval state smoother is presented by Anderson & Moore (1979) but a more efficient state smoother has recently been developed by de Jong (1988, 1989) and Kohn & Ansley (1989). This paper presents a smoother that evaluates the minimum mean squared error estimator of the disturbance vector ε_t and the associated mean squared error matrices. The algorithm is computationally efficient, it does not use a state smoother and it has some powerful applications, including an EM algorithm for estimating covariances of a state space model and a smoother for the state vector.

2. THE DISTURBANCE SMOOTHER

2·1. *The algorithm*

The disturbance smoother evaluates the minimum mean squared error estimator of the disturbance vector, that is the smoothed residual $\tilde{\varepsilon}_t = E(\varepsilon_t \mid y)$, where $y = (y_1', \ldots, y_n')'$, and the corresponding mean squared error matrices. We proceed by presenting the disturbance smoother, under the assumption that the parameter vector δ is known, in order to obtain expressions for $u_t = E(\varepsilon_t \mid y, \delta)$ and the related mean squared error matrices $\mathrm{MSE}(u_t, u_s)$, for $t = 1, \ldots, n$ and $s = 1, \ldots, t$. The special case $H_t G_t' = 0$, for $t = 1, \ldots, n$, is discussed in more detail in § 2·2. The disturbance smoother is generalized to allow for an unknown vector δ in § 2·3.

Assume that the parameter vector δ is known, then the smoothed residual vectors u_t and their contemporaneous and lagged mean squared error matrices are obtained by first applying the Kalman filter and storing the quantities v_t, F_t^{-1} and K_t, for $t = 1, \ldots, n$, and, subsequently, applying the backwards disturbance smoother as given by

$$\begin{pmatrix} r_{t-1} \\ u_t \end{pmatrix} = \begin{pmatrix} Z_t & G_t \\ T_t & H_t \end{pmatrix}' \begin{pmatrix} e_t \\ r_t \end{pmatrix} \quad (t = n, \ldots, 1), \tag{2·1}$$

$$\begin{pmatrix} N_{t-1} & C_t^{*\prime} \\ C_t^* & C_t \end{pmatrix} = \begin{pmatrix} Z_t & G_t \\ T_t & H_t \end{pmatrix}' \begin{pmatrix} D_t & -K_t' N_t \\ -N_t K_t & N_t \end{pmatrix} \begin{pmatrix} Z_t & G_t \\ T_t & H_t \end{pmatrix} \quad (t = n, \ldots, 1), \tag{2·2}$$

where $e_t = F_t^{-1} v_t - K_t' r_t$ and $D_t = F_t^{-1} + K_t' N_t K_t$. The vector r_t and the matrix N_t are dummy quantities to generate the backwards recursion. The disturbance smoother is initialized with $r_n = 0$ and $N_n = 0$. The smoothed residuals u_t are obtained directly from the disturbance smoother and the corresponding mean squared error matrices are given by

$$\mathrm{MSE}(u_t) = \sigma^2 (I - C_t) \quad (t = n, \ldots, 0), \tag{2·3}$$

$$\mathrm{MSE}(u_t, u_s) = -\sigma^2 C_t^* L_{t,s+1} M_s \quad (s = t-1, \ldots, 0), \tag{2·4}$$

where the matrix $L_{t,j}$ is given by the product $L_{t-1}L_{t-2}\ldots L_j$, for $j=1,\ldots,t-1$, and the matrix $L_{j,j}$ is the identity matrix, for $j=1,\ldots,n$. Finally, we define $u_0=H'_0r_0$, $C_0=H'_0N_0H_0$ and $M_0=H_0$. The proof of the disturbance smoother is given in Appendix 2.

The computational costs of the disturbance smoother are modest. For many time series models, the disturbance smoother requires a number of flops that is about the same as for the Kalman filter. Thus, the disturbance smoother is more efficient than existing state smoothers as they require many more flops than the Kalman filter. Table 1 gives the computational costs of the Kalman filter, the disturbance smoother and a state smoother for a range of unobserved components time series models.

Table 1. *Computational costs*

Model	Dimensions y_t	α_t	ε_t	Number of flops: Kalman filter	Disturb. smooth.	State smooth.	Storage: Disturb. smooth.	State smooth.
Local level	1	1	2	3	5	8	3	5
Local linear trend	1	2	3	10	10	32	4	10
Quarterly basic	1	5	4	40	37	332	7	37
Monthly basic	1	13	4	208	197	4916	15	197
Multivariate	5	5	10	400	300	575	55	85

Kalman filter includes the recursions for a_t and P_t; disturbance smoother includes u_t and C_t; state smoother is described by de Jong (1989). Unobserved components time series models are described by Harvey (1989, p. 510). Basic model is discussed in § 3·1; multivariate model refers to a local level model containing a vector of five time series. Number of flops and storage are counted per observation.

2·2. *A special case*

For many time series models cast in the state space form (1·1)–(1·2), the restriction $H_tG'_t=0$ for $t=1,\ldots,n$ holds, which implies that the measurement and the system equations are independent. Under these circumstances, the smoothed estimators of the disturbances $G_t\varepsilon_t$ and $H_t\varepsilon_t$ are defined by $G_tu_t=G_tG'_te_t$ and $H_tu_t=H_tH'_tr_t$ $(t=1,\ldots,n)$, respectively, and they are evaluated by the backwards recursion

$$e_t=F_t^{-1}v_t-K'_tr_t,\quad r_{t-1}=Z'_te_t+T'_tr_t\quad (t=n,\ldots,1), \tag{2·5}$$

with $r_n=0$. Thus, in this case the dummy quantities e_t and r_t have a special interpretation: they are the scaled estimators $E(G_t\varepsilon_t\,|\,y,\delta)$ and $E(H_t\varepsilon_t\,|\,y,\delta)$ corresponding to the disturbances of the measurement and the system equations, respectively.

Also, the dummy matrices D_t and N_t are directly related to the mean squared error matrices of G_tu_t and H_tu_t, respectively, since

$$\text{MSE}(G_tu_t)=\sigma^2G_tG'_t-\sigma^2G_tG'_tD_tG_tG'_t, \tag{2·6a}$$

$$\text{MSE}(H_tu_t)=\sigma^2H_tH'_t-\sigma^2H_tH'_tN_tH_tH'_t\quad (t=1,\ldots,n), \tag{2·6b}$$

$$\text{MSE}(G_tu_t,\,H_tu_t)=-\sigma^2G_tG'_tK'_tN_tH_tH'_t. \tag{2·6c}$$

These mean squared error expressions are evaluated by the backwards recursion

$$D_t=F_t^{-1}+K'_tN_tK_t,\quad N_{t-1}=Z'_tD_tZ_t-T'_tN_tK_tZ_t-Z'_tK'_tN_tT_t+T'_tN_tT_t\quad (t=n,\ldots,1),$$

with $N_n = 0$. Finally, the reduced expressions for the lagged mean squared error matrices are

$$\mathrm{MSE}(G_t u_t, G_s u_s) = -\sigma^2 G_t G_t'(K_t' N_t T_t - D_t Z_t) L_{t,s+1} K_s G_s G_s', \tag{2·7a}$$

$$\mathrm{MSE}(H_t u_t, H_s u_s) = -\sigma^2 H_t H_t' N_t (T_t - K_t Z_t) L_{t,s+1} H_s H_s' \quad (t = 1, \ldots n), \tag{2·7b}$$

$$\mathrm{MSE}(G_t u_t, H_s u_s) = \sigma^2 G_t G_t'(K_t' N_t T_t - D_t Z_t) L_{t,s+1} H_s H_s' \quad (s = 1, \ldots, t-1), \tag{2·7c}$$

$$\mathrm{MSE}(H_t u_t, G_s u_s) = \sigma^2 H_t H_t' N_t (T_t - K_t Z_t) L_{t,s+1} K_s G_s G_s', \tag{2·7d}$$

which can be evaluated by using the backwards recursions for D_t and N_t as well. These results follow directly when the restriction $H_t G_t' = 0$ $(t = 1, \ldots, n)$ is applied to the results of § 2·1.

2·3. *The augmented disturbance smoother*

To allow for the unknown parameter vector δ, the disturbance smoother must be augmented in a similar way as the Kalman filter is augmented in Appendix 1. The augmented Kalman filter produces the minimum mean squared error estimator of δ based on y, that is d, together with its mean squared error matrix $\sigma^2 S^{-1}$. To incorporate this estimator in the disturbance smoother, we replace the recursion (2·1) by

$$\begin{pmatrix} R_{t-1} \\ U_t \end{pmatrix} = \begin{pmatrix} Z_t & G_t \\ T_t & H_t \end{pmatrix}' \begin{pmatrix} E_t \\ R_t \end{pmatrix} \quad (t = n, \ldots, 1), \tag{2·8}$$

with $R_n = 0$, $E_t = F_t^{-1} V_t - K_t' R_t$ and V_t as defined in Appendix 1. Now, the smoothed estimator of the disturbance vector with its mean squared error matrices are

$$\tilde{\varepsilon}_t = E(\varepsilon_t \mid y) = u_t^\dagger + U_t^\dagger d, \tag{2·9}$$

$$\mathrm{MSE}(\tilde{\varepsilon}_t) = \sigma^2 (I - C_t + U_t^\dagger S^{-1} U_t^{\dagger\prime}) \quad (t = n, \ldots, 1), \tag{2·10}$$

$$\mathrm{MSE}(\tilde{\varepsilon}_t, \tilde{\varepsilon}_j) = \sigma^2 (-C_t^* L_{t,j+1} M_j + U_t^\dagger S^{-1} U_j^{\dagger\prime}) \quad (j = t-1, \ldots, 1), \tag{2·11}$$

where U_t is partitioned as $(u_t^\dagger, U_t^\dagger)$. These results are valid when unknown δ is supposed to be fixed or diffuse, since the inference results with respect to δ are the same for both conditions (de Jong, 1991). The derivation of the smoothed residuals, for δ unknown, is given in Appendix 3.

3. APPLICATIONS

3·1. *Auxiliary residuals for unobserved components time series models*

This section discusses some practical implications of the results derived in § 2. All applications assume δ known since it is straightforward to generalize the results for δ unknown.

Auxiliary residuals are defined as the estimators of disturbances associated with unobserved components. Harvey & Koopman (1993) show that these residuals often yield information which is less apparent from the one-step ahead prediction errors. For example, consider the basic structural time series model

$$y_t = \mu_t + \gamma_t + w_t, \quad w_t \sim \mathrm{NID}(0, \sigma^2) \quad (t = 1, \ldots, n),$$

where the unobserved components μ_t, γ_t and w_t are interpreted as trend, seasonal and irregular, respectively. The stochastic processes of the components are given by

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \quad \eta_t \sim \text{NID}(0, q_\eta \sigma^2),$$

$$\beta_t = \beta_{t-1} + \zeta_t, \quad \zeta_t \sim \text{NID}(0, q_\zeta \sigma^2),$$

$$\sum_{j=0}^{s-1} \gamma_{t-j} = \omega_t, \quad \omega_t \sim \text{NID}(0, q_\omega \sigma^2), \quad (t = 1, \ldots, n),$$

where s is the number of seasons, β_t is interpreted as the slope of the trend and q_η, q_ζ and q_ω are signal-to-noise ratios (Harvey, 1989, p. 33). The basic structural time series model can be placed in a state space model with time-invariant system matrices Z_t, T_t, G_t and H_t, state vector $\alpha_t = (\mu_t, \beta_t, \gamma_t, \gamma_{t-1}, \gamma_{t-2})'$ and disturbance vectors $G\varepsilon_t = w_t$ and $H\varepsilon_t = (\eta_{t+1}, \zeta_{t+1}, \omega_{t+1})'$, where the restriction $HG' = 0$ holds.

The auxiliary residuals are the smoothed estimators of $G\varepsilon_t$ and $H\varepsilon_t$ standardized by dividing by the square root of their mean squared errors, that is

$$\tilde{w}_t = e_t, \quad \text{MSE}(\tilde{w}_t) = \sigma^2(1 - D_t), \tag{3·1a}$$

$$\tilde{\eta}_t = q_\eta r^1_{t-1}, \quad \text{MSE}(\tilde{\eta}_t) = q_\eta \sigma^2(1 - q_\eta N^{1,1}_{t-1}), \tag{3·1b}$$

$$\tilde{\zeta}_t = q_\zeta r^2_{t-1}, \quad \text{MSE}(\tilde{\zeta}_t) = q_\zeta \sigma^2(1 - q_\zeta N^{2,2}_{t-1}), \tag{3·1c}$$

$$\tilde{\omega}_t = q_\omega r^3_{t-1}, \quad \text{MSE}(\tilde{\omega}_t) = q_\omega \sigma^2(1 - q_\omega N^{3,3}_{t-1}) \quad (t = 1, \ldots, n), \tag{3·1d}$$

where r^i_t is the ith element of vector r_t and $N^{i,j}_t$ is the (i, j)th element of matrix N_t. These residuals are serially correlated. The theoretical auto- and cross-correlation functions of the residuals can be obtained via the disturbance smoother since it includes expressions for the lagged mean squared error matrices; see § 2·2. The plots of the auxiliary residuals together with diagnostic tests for normality and kurtosis, corrected to allow for the implied serial correlation, provide useful information to detect time series irregularities (Harvey & Koopman, 1993). For example, a level shift is detected from the auxiliary residual related to η_t and an additive outlier appears in the residual corresponding to the irregular w_t.

3·2. *A smoother for the state vector*

If mean squared error values are not required, the results in § 2 lead to an efficient algorithm for calculating the smoothed state vector, that is the minimum mean squared error estimator of the state vector based on y and δ, as denoted by $\tilde{\alpha}_t = E(\alpha_t | y, \delta)$. It follows directly from the state equation of the state space model that the smoothed state can be evaluated recursively by the state smoother as given by

$$\tilde{\alpha}_{t+1} = T_t \tilde{\alpha}_t + W_t \beta + H_t u_t \quad (t = 1, \ldots, n), \tag{3·2}$$

which is started off with $\tilde{\alpha}_1 = W_0 \beta + H_0 u_0$. This method of state smoothing consists of three steps: (i) the Kalman filter stores v_t, F_t^{-1} and K_t for $t = 1, \ldots, n$; (ii) the disturbance smoother partially overwrites the Kalman filter storage by u_t for $t = n, \ldots, 1$; (iii) the state smoother (3·2) computes the smoothed state for $t = 1, \ldots, n$. The mean squared error quantities of $\tilde{\alpha}_t$ cannot be evaluated efficiently by this approach and, therefore, the state smoother (3·2) does not require the disturbance smoother recursion for the matrices N_t, C_t and C^*_t.

The original fixed interval state smoother is presented by Anderson & Moore (1979, p. 189, eqn. (4.5)) as

$$\tilde{\alpha}_t = a_t^* + P_t^*(\tilde{\alpha}_{t+1} - a_{t+1}) \quad (t = n-1, \ldots, 1), \tag{3·3}$$

where $a_t^* = a_t + P_t Z_t' F_t^{-1} v_t$ and $P_t^* = P_t L_t' P_{t+1}^{-1}$. This recursion is initialized by $\tilde{\alpha}_n = a_n^*$. The state smoother of de Jong (1988, 1989) and Kohn & Ansley (1989) is given by

$$\tilde{\alpha}_t = a_t + P_t r_{t-1} \quad (t = n, \ldots, 1), \tag{3·4}$$

where r_t is evaluated as in the disturbance smoother. The state smoothers (3·3) and (3·4) can be applied after the Kalman filter has stored v_t, F_t^{-1}, K_t, a_t and P_t for $t = 1, \ldots, n$. The smoothing methods based on (3·3) and (3·4) also provide recursions to evaluate the mean squared error matrices of the smoothed state vectors.

Table 2 reports the required number of flops for the original fixed interval state smoothers (3·2), (3·3) and (3·4) for some time-point t. The time series models listed in Table 2 fall into the class of unobserved components time series models (Harvey, 1989, p. 510). The state smoother (3·2) is more efficient for this range of time series models. The same claim can be made in general for time series models which have system matrices consisting of many zero and unity values.

Table 2. *Number of flops for state smoothers*

	Dimensions			Number of flops			
Model	y_t	α_t	ε_t	(2·1)	(3·2)	(3·3)	(3·4)
Local level	1	1	2	2	1	5	1
Local linear trend	1	2	3	3	2	26	4
Quarterly basic	1	5	4	6	3	406	25
Monthly basic	1	13	4	14	3	6774	169
Multivariate	5	5	10	50	5	451	25

Unobserved components time series models are described by Harvey (1989, p. 510). Basic model is discussed in § 3·1; multivariate model refers to a local level model containing a vector of five time series. Smoothers (3·2) and (3·4) require the recursion (2·1). Number of flops is counted per observation.

3·3. *EM algorithm for estimating covariances in state space models*

Maximum likelihood estimation of unknown elements in the system matrices of the Gaussian state space model is normally carried out by numerically maximizing the log likelihood. An alternative approach, based on the EM algorithm, is proposed by Shumway & Stoffer (1982) and Watson & Engle (1983). The computational cost of this EM is rather high since it requires a smoother for the state vector and the corresponding mean squared error matrices. Under the circumstances where only unknown elements are found in the covariance matrices of the state space model, an EM algorithm can be used that only requires the disturbance smoother without C_t^*, so saving a considerable computational effort; see Table 1.

Consider the Gaussian state space model with time invariant system matrices $G_t = G$, $H_t = H$ and $HG' = 0$. Let λ be the stack of unknown parameters in the covariance matrices $\Omega_G = GG'$ and $\Omega_H = HH'$ such that $\Omega_G = \Omega_G(\lambda)$ and $\Omega_H = \Omega_H(\lambda)$. Then, given an initial estimate of the parameter vector λ, say λ^*, the EM step gives a new set of covariance matrices $\Omega_G(\lambda^\dagger)$ and $\Omega_H(\lambda^\dagger)$ which always increases the log likelihood value. The new covariances are

$$\Omega_G(\lambda^\dagger) = \Omega_G(\lambda^*) + \Omega_G(\lambda^*)\Theta_e\Omega_G(\lambda^*), \tag{3·5}$$

$$\Omega_H(\lambda^\dagger) = \Omega_H(\lambda^*) + \Omega_H(\lambda^*)\Theta_r\Omega_H(\lambda^*) \quad (t=1, \ldots, n), \tag{3·6}$$

where

$$\Theta_e = n^{-1}\sum_{t=1}^{n}\sigma^{-2}e_t e_t' - D_t, \quad \Theta_r = n^{-1}\sum_{t=1}^{n}\sigma^{-2}r_t r_t' - N_t. \tag{3·7}$$

The smoothed quantities e_t, D_t, r_t and N_t are computed by the disturbance smoother with parameter vector $\lambda = \lambda^*$. The proof of the EM step for the covariances is given in Appendix 4. The maximum likelihood estimates of the unknown elements in the covariance matrices are obtained by repeatedly applying EM steps until the increase of the log likelihood is negligible. The EM algorithm can become rather slow in convergence when it comes close to the optimum. Under these circumstances, one may switch to another numerical optimization procedure. Such a method may require the score vector with respect to λ. Koopman & Shephard (1992) have developed a computationally efficient method to calculate the exact score whch is based on the disturbance smoother and the matrices Θ_e and Θ_r.

Acknowledgements

I am indebted to Andrew Harvey, Neil Shephard, Piet de Jong and Jim Durbin for their support and many discussions. I also thank the editor and two anonymous referees for their useful comments on an earlier draft. All remaining errors are mine.

Appendix 1

The augmented Kalman filter

When the parameter vector δ is unknown, an estimator for δ is constructed via a partially adjusted Kalman filter. We follow de Jong (1991) by augmenting the Kalman filter by a matrix A_t and by re-defining the minimum mean squared error estimator of the state vector as $a_t = a_t^\dagger + A_t^\dagger\delta$, where $A_t = (a_t^\dagger, A_t^\dagger)$. The vector $a_t^\dagger$ corresponds to the known part of the initial state vector, that is $W_0 b$, whereas $A_t^\dagger$ allows for the correction when δ is unknown. Similarly, the one-step ahead prediction error becomes $v_t = v_t^\dagger + V_t^\dagger\delta$ and the matrix $V_t = (v_t^\dagger, V_t^\dagger)$ corresponds to A_t. Thus, the equations for v_t and a_t in the Kalman filter are replaced by

$$V_t = (y_t, 0) - Z_t A_t - X_t(b, B), \quad A_{t+1} = T_t A_t + W_t(b, B) + K_t V_t \quad (t=1, \ldots, n),$$

where 0 is a matrix of zeros and $A_1 = W_0(b, B)$. In addition, the augmented Kalman filter contains the matrix recursion $S_t^\dagger = S_{t-1}^\dagger + V_t' F_t^{-1} V_t$ $(t=1, \ldots, n)$, where $S_t^\dagger$ is partitioned as

$$S_t^\dagger = \begin{pmatrix} q_t & s_t' \\ s_t & S_t \end{pmatrix},$$

and $S_0^\dagger$ is a zero matrix. Assuming that the parameter vector δ is fixed but unknown, the estimator of δ at time t, that is the minimum mean squared error estimator based on Y_t, and its mean squared error matrix are given by

$$d_t = S_t^{-1}s_t, \quad \mathrm{MSE}(d_t) = \sigma^2 S_t^{-1} \quad (t=1, \ldots, n),$$

respectively. It follows that the estimator of the state vector at time t, based on Y_{t-1}, is given by $a_t^\dagger + A_t^\dagger d_{t-1}$ with its mean squared error matrix $\sigma^2(P_t + A_t^\dagger S_{t-1}^{-1} A_t^{\dagger\prime})$, for $t=1,\ldots,n$. When the vector δ is regarded as diffuse, that is $\delta \sim (\mu, \sigma^2\Omega)$ where Ω^{-1} converges to zero in the Euclidean norm, these results remain the same as with δ fixed (de Jong, 1991).

APPENDIX 2

Proof of disturbance smoother

The proof of the disturbance smoother starts off with the definition of the one-step ahead prediction error of the state $x_t = \alpha_t - a_t$, for $t=1,\ldots,n$. It follows from the Kalman filter that

$$x_{t+1} = L_t x_t + M_t \varepsilon_t, \quad \operatorname{cov}(x_{t+1}) = \sigma^2 P_{t+1},$$

$$v_t = Z_t x_t + G_t \varepsilon_t, \quad \operatorname{cov}(v_t) = \sigma^2 F_t \quad (t=1,\ldots,n),$$

Thus, the one-step ahead prediction error can be written as

$$v_j = Z_j L_{j,1} x_1 + Z_j \sum_{t=1}^{j-1} L_{j,t+1} M_t \varepsilon_t + G_j \varepsilon_j \quad (j=1,\ldots,n),$$

such that the cross-covariances between the disturbances and the one-step ahead prediction errors are

$$\operatorname{cov}(\varepsilon_t, v_j) = \begin{cases} 0 & (j=1,\ldots,t-1), \\ \sigma^2 G_t' & (j=t), \\ \sigma^2 (Z_j L_{j,t+1} M_t)' & (j=t+1,\ldots,n), \end{cases}$$

for $t=1,\ldots,n$. Furthermore, $\operatorname{cov}(\varepsilon_0, v_j) = \sigma^2(Z_j L_{j,1} H_0)'$ for $j=1,\ldots,n$.

Because the stack of one-step ahead prediction errors $v=(v_1',\ldots,v_n')'$ is a linear transformation of $(y',\delta')'$, it follows that $u_t = E(\varepsilon_t \mid y,\delta) = E(\varepsilon_t \mid v)$. Also, the one-step ahead prediction errors are uncorrelated so that $\operatorname{cov}(v) = \sigma^2 F$ is the block diagonal matrix $\sigma^2 \operatorname{diag}(F_1,\ldots,F_n)$. Applying standard results of linear estimation,

$$\begin{aligned} u_t = \operatorname{cov}(\varepsilon_t, v)\operatorname{cov}(v)^{-1} v &= \sigma^{-2} \sum_{j=1}^{n} \operatorname{cov}(\varepsilon_t, v_j) F_j^{-1} v_j \\ &= G_t' F_t^{-1} v_t + \sum_{j=t+1}^{n} (Z_j L_{j,t+1} M_t)' F_j^{-1} v_j \\ &= G_t' F_t^{-1} v_t + M_t' r_t, \end{aligned} \tag{A2·1}$$

and $r_t = \Sigma (Z_j L_{j,t+1})' F_j^{-1} v_j$, where the summation is over the range $j=t+1,\ldots,n$, is evaluated by the backwards recursion

$$r_{t-1} = Z_t' F_t^{-1} v_t + L_t' r_t, \tag{A2·2}$$

for $t=n,\ldots,1$. It follows that $r_n = 0$. In a similar way it can be shown that $u_0 = H_0' r_0$.

We derive the mean squared error matrix of u_t by

$$\begin{aligned} \text{MSE}(u_t) = \operatorname{cov}(\varepsilon_t, \varepsilon_t - u_t) &= \operatorname{cov}(\varepsilon_t, \varepsilon_t - G_t' F_t^{-1} v_t) - \operatorname{cov}(\varepsilon_t, r_t) M_t \\ &= \sigma^2 (I - G_t' F_t^{-1} G_t) - \left\{ \sum_{j=t+1}^{n} \operatorname{cov}(\varepsilon_t, v_j) F_j^{-1} (Z_j L_{j,t+1}) \right\} M_t \\ &= \sigma^2 (I - G_t' F_t^{-1} G_t - M_t' N_t M_t), \end{aligned} \tag{A2·3}$$

where the matrix $N_t = \Sigma (Z_j L_{j,t+1})' F_j^{-1} (Z_j L_{j,t+1})$, with the summation over the range $j=t+1,\ldots,n$, is evaluated by the backwards recursion

$$N_{t-1} = Z_t' F_t^{-1} Z_t + L_t' N_t L_t, \tag{A2·4}$$

for $t = n, \ldots, 1$. The recursion (A2·4) is initialized by $N_n = 0$. Via the same steps, it follows that $\text{MSE}(u_0) = \sigma^2(I - H_0' N_0 H_0)$.

Finally, we obtain $\text{MSE}(u_t, u_s)$, for $s = 1, \ldots, t-1$, by

$$\begin{aligned}
\text{MSE}(u_t, u_s) &= \text{cov}(\varepsilon_t, \varepsilon_s - u_s) = \text{cov}(\varepsilon_t, \varepsilon_s - G_s' F_s^{-1} v_s) - \text{cov}(\varepsilon_t, r_s) M_s \\
&= -\sum_{j=s+1}^{n} \text{cov}(\varepsilon_t, v_j) F_j^{-1}(Z_j L_{j,s+1}) M_s \\
&= -\text{cov}(\varepsilon_t, v_t) F_t^{-1}(Z_t L_{t,s+1}) M_s - \sum_{j=t+1}^{n} \text{cov}(\varepsilon_t, v_j) F_j^{-1}(Z_j L_{j,s+1}) M_s \\
&= -\sigma^2(G_t' F_t^{-1} Z_t + M_t' N_t L_t) L_{t,s+1} M_s, \qquad \text{(A2·5)}
\end{aligned}$$

and, in a similar way, $\text{MSE}(u_t, u_0) = -\sigma^2(G_t' F_t^{-1} Z_t + M_t' N_t L_t) L_{t,1} H_0$ for $t = 1, \ldots, n$.

By exploiting the definitions of the matrices L_t and M_t and after some minor manipulations with the equations (A2·1)–(A2·5), the disturbance smoother appears in the presented matrix form (2·1) and (2·2).

Appendix 3

Derivation of the smoothed disturbances when δ is unknown

This appendix derives equations for the smoothed disturbances and the corresponding mean squared error matrices assuming that the parameter vector δ is unknown. In Appendix 1, the prediction error v_t is defined as $v_t = v_t^\dagger + V_t^\dagger \delta$. By substituting this definition into (A2·1) and (A2·2), the augmented disturbance smoother expressions for $U_t = (u_t^\dagger, U_t^\dagger)$ and $R_t = (r_t^\dagger, R_t^\dagger)$ follow straightforwardly when, additionally, we define $u_t = u_t^\dagger + U_t^\dagger \delta$ and $r_t = r_t^\dagger + R_t^\dagger \delta$. Now,

$$\tilde{\varepsilon}_t = E(\varepsilon_t | y) = E\{E(\varepsilon_t | y, \delta) | y\} = E(u_t^\dagger + U_t^\dagger \delta | y) = u_t^\dagger + U_t^\dagger E(\delta | y) \quad (t = n, \ldots, 1), \qquad \text{(A3·1)}$$

$$\begin{aligned}
\text{MSE}(\tilde{\varepsilon}_t, \tilde{\varepsilon}_j) &= \text{MSE}(\varepsilon_t, \varepsilon_j | y) = \text{cov}(\varepsilon_t - u_t + u_t - \tilde{\varepsilon}_t, \varepsilon_j - u_j + u_j - \tilde{\varepsilon}_j) \\
&= \text{MSE}(u_t, u_j) + \text{cov}(U_t^\dagger(\delta - d_n), U_j^\dagger(\delta - d_n)) \\
&= \text{MSE}(u_t, u_j) + U_t^\dagger \text{MSE}(d_n) U_j^{\dagger\prime} \quad (j = t, \ldots, 1). \qquad \text{(A3·2)}
\end{aligned}$$

The smoothed estimators and the mean squared error matrices can be evaluated by the augmented disturbance smoother as presented in § 2·3.

Appendix 4

Proof of the EM step for estimating covariances

This appendix derives the EM step for the covariances in the Gaussian state space model (1·1)–(1·2). The system matrices G_t and H_t are time-invariant and $HG' = 0$. The covariances are defined by $\Omega_G = GG'$ and $\Omega_H = HH'$ and the unknown elements are stacked into the vector λ such that $\Omega_G = \Omega_G(\lambda)$ and $\Omega_H = \Omega_H(\lambda)$. Following Shumway & Stoffer (1982), the joint log likelihood of the complete data-set $\alpha = (\alpha_0', \alpha_1', \ldots, \alpha_{n+1}')'$ and $y = (y_1', \ldots, y_n')'$ is, apart from constants and assuming δ known,

$$-\tfrac{1}{2}\sum_{t=1}^{n} \{\log|\sigma^2 \Omega_G| + \sigma^2 \,\text{tr}\,(\Omega_G^{-1} G \varepsilon_t \varepsilon_t' G')\} - \tfrac{1}{2}\sum_{t=1}^{n} \{\log|\sigma^2 \Omega_H| + \sigma^{-2}\,\text{tr}\,(\Omega_H^{-1} H \varepsilon_t \varepsilon_t' H')\}, \qquad \text{(A4·1)}$$

where $G\varepsilon_t = y_t - Z_t \alpha_t - X_t \beta$ and $H\varepsilon_t = \alpha_{t+1} - T_t \alpha_t - W_t \beta$ for $t = 1, \ldots, n$. The expectation of (A4·1), conditional on the observed data y and relative to the likelihood containing a specific value for

λ, say λ^*, is given by

$$-\tfrac{1}{2}\sum_{t=1}^{n}(\log|\sigma^2\Omega_G|+\sigma^{-2}\operatorname{tr}[\Omega_G^{-1}\{\Omega_G(\lambda^*)+\Omega_G(\lambda^*)(e_te_t'-\sigma^2D_t)\Omega_G(\lambda^*)\}])$$

$$-\tfrac{1}{2}\sum_{t=1}^{n}(\log|\sigma^2\Omega_H|+\sigma^{-2}\operatorname{tr}[\Omega_H^{-1}\{\Omega_H(\lambda^*)+\Omega_H(\lambda^*)(r_tr_t'-\sigma^2N_t)\Omega_H(\lambda^*)\}]). \qquad (A4{\cdot}2)$$

Now it is easy to see that the EM step, as given in § 3·3, produces the new estimates for Ω_G and Ω_H that maximize (A4·2).

REFERENCES

ANDERSON, B. D. O. & MOORE, J. B. (1979). *Optimal Filtering*. Englewood Cliffs, New Jersey: Prentice Hall.
DE JONG, P. (1988). A cross-validation filter for time series models. *Biometrika* **75**, 594–600.
DE JONG, P. (1989). Smoothing and interpolation with the state space model. *J. Am. Statist. Assoc.* **84**, 1085–8.
DE JONG, P. (1991). The diffuse Kalman filter. *Ann. Statist.* **19**, 1073–83.
HARVEY, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.
HARVEY, A. C. & KOOPMAN, S. J. (1993). Diagnostic checking of unobserved components time series models. *J. Bus. Econ. Statist.* To appear.
KOHN, R. & ANSLEY, C. F. (1989). A fast algorithm for signal extraction, influence and cross-validation in state space models. *Biometrika* **76**, 65–79.
KOOPMAN, S. J. & SHEPHARD, N. G. (1992). Exact score for time series models in state space form. *Biometrika* **79**, 823–6.
SHUMWAY, R. H. & STOFFER, D. S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *J. Time Ser. Anal.* **3**, 253–64.
WATSON, M. W. & ENGLE, R. F. (1983). Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression. *J. Econometrics* **23**, 385–400.

[*Received February* 1991. *Revised May* 1992]