FAST FILTERING AND SMOOTHING FOR MULTIVARIATE STATE SPACE MODELS

By S. J. Koopman and J. Durbin

Free University Amsterdam
London School of Economics and Political Science

First version received March 1998

Abstract. This paper investigates a new approach to diffuse filtering and smoothing for multivariate state space models. The standard approach treats the observations as vectors, while our approach treats each element of the observational vector individually. This strategy leads to computationally efficient methods for multivariate filtering and smoothing. Also, the treatment of the diffuse initial state vector in multivariate models is much simpler than in existing methods. The paper presents details of relevant algorithms for filtering, prediction and smoothing. Proofs are provided. Three examples of multivariate models in statistics and economics are presented for which the new approach is particularly relevant.

Keywords. Diffuse initialization; Kalman filter; multivariate models; smoothing; state space; time series; vector cubic splines.

1. INTRODUCTION

In the standard multivariate linear state space model, the observation vector y_t depends linearly on an unobserved state vector α_t, which develops over time as a first-order vector autoregression for $t = 1, \ldots, n$. In this paper we consider filtering and smoothing for this model. The object of filtering is to calculate the mean and error variance matrix of α_t given $y_1, \ldots, y_{t-1}$ and the object of smoothing is to calculate the mean and error variance matrix of α_t given $y_1, \ldots, y_n$. Analysis based on these models is important in many areas and particularly in applied time series analysis. For a general treatment of state space models for time series analysis see Harvey (1989) and for an application to a particular problem of public importance together with a published discussion of the merits of these models see Harvey and Durbin (1986).

The conventional approach to filtering and smoothing for these models is based on considering the contribution of the entire observational vector at each successive time point. The basic idea of this paper is to introduce the elements of the observational vectors one at a time into the filtering and smoothing processes. In effect, we convert the original multivariate series into a univariate series and analyse the data in univariate form. Although the concept is simple, the improvement in computational efficiency is dramatic for models of more

0143-9782/00/04 281–296 JOURNAL OF TIME SERIES ANALYSIS Vol. 21, No. 3

than a modest degree of complexity. The advantage is particularly strong for the treatment of initialization by diffuse priors.

The idea of decomposing the observational vectors into sub-vectors for the improvement of computational efficiency in Kalman filtering was suggested by Anderson and Moore (1979, Section 6.4) under the name sequential processing. Fahrmeir and Tutz (1996, Section 8.4) discuss a similar strategy for longitudinal models. However, both contributions assume that the initial conditions are known and they do not deal with diffuse initialization and parameter estimation which are major concerns in this paper. Ansley and Kohn (1990, Section 4) also mention the univariate approach, which they use in their treatment of diffuse filtering only. In this paper we give a full treatment of filtering and smoothing for state space models (non-diffuse and diffuse) that is simple and easy to implement on a computer.

Section 2 presents the multivariate linear Gaussian state space model and sets out the standard Kalman filter recursions in a form that is suitable for later work in the paper. A general form of the partially diffuse initial state vector is considered in which some elements of the state vector at the initial time point have finite variances while others have infinite variances. In Section 3 the model is written in univariate form, first for the case where the observation error matrix is diagonal and second for the case where the matrix is an arbitrary positive semi-definite matrix. Section 4 begins by deriving the Kalman filtering recursion for the main part of the univariate seres and goes on to consider the special features of the recursions that are needed to handle the time points at the beginning of the series that are directly affected by the diffuse initialization. In Section 5, recursions are given for the state and disturbance smoothing, first for the main part of the series and then for the part at the beginning that is affected by the diffuse initialization. Maximum likelihood estimation of parameters is considered in Section 6. Three examples of multivariate models in state space form are given in Section 7; the saving in computing can be dramatic in some cases, as is shown. Section 8 concludes.

2. REVIEW OF STANDARD STATE SPACE METHODS

2.1. *State space model*

The multivariate Gaussian linear state space model is given by

$$\begin{aligned} y_t &= Z_t\alpha_t + \varepsilon_t \qquad \varepsilon_t \sim \mathrm{N}(0,\ H_t) \\ \alpha_{t+1} &= T_t\alpha_t + R_t\eta_t \qquad \eta_t \sim \mathrm{N}(0,\ Q_t), \quad t = 1, \ldots, n \end{aligned} \tag{1}$$

where y_t is the $p_t \times 1$ vector of observations, α_t is the $m \times 1$ state vector and ε_t is the $p_t \times 1$ vector of disturbances. The state vector follows a Markov process with $q \times 1$ disturbance vector η_t. The equation for y_t is called the observation equation and the equation for α_{t+1} is referred to as the state equation. The normally and independently distributed disturbance vectors ε_t and η_t are

mutually uncorrelated. The initial state vector is assumed to be normally distributed with mean vector a and variance matrix P, i.e. $\alpha_1 \sim \mathrm{N}(a, P)$. The system matrices Z_t, H_t, T_t, R_t and Q_t, with appropriate dimensions, are fixed matrices. The state space model (1) is said to be time invariant when the system matrices are constant over time index t. In many practical situations the state space model can be set up as time invariant.

When the state vector contains non-stationary components or regression effects, elements of the initial state vector α_1 may require a diffuse prior. We therefore assume that the distribution of α_1 has the general form

$$\alpha_1 \sim \mathrm{N}(a, P) \qquad P = \kappa P_\infty + P_* \qquad \kappa > 0 \tag{2}$$

where vector a and matrices P_∞ and P_* are fixed and known and where we shall in due course let $\kappa \to \infty$. The matrix P_∞ is typically diagonal and when a diagonal element of P_∞ is non-zero the corresponding row and column of P_* are not relevant.

2.2. *Kalman filter*

The Kalman filter recursions evaluate the mean of the state vector α_{t+1} conditional on the observations $Y_t = \{y_1, \ldots, y_t\}$ and its error variance matrix, i.e. $a_{t+1} = E(\alpha_{t+1}|Y_t)$ and $P_{t+1} = \mathrm{var}(\alpha_{t+1}|Y_t)$, for $t = 1, \ldots, n$. The Kalman filter for the state space model (1) and (2) with κ given can be written in the form

$$\begin{gathered} v_t = y_t - Z_t a_t \qquad F_t = Z_t P_t Z_t' + H_t \\ K_t = P_t Z_t' \\ a_{t+1} = T_t(a_t + K_t F_t^{-1} v_t) \qquad P_{t+1} = T_t(P_t - K_t F_t^{-1} K_t') T_t' + R_t Q_t R_t' \end{gathered} \tag{3}$$

for $t = 1, \ldots, n$. The one-step-ahead prediction error is $v_t = y_t - E(y_t|Y_{t-1})$ with variance matrix $F_t = \mathrm{var}(y_t|Y_{t-1}) = \mathrm{var}(v_t)$. The matrix K_t is the covariance matrix $\mathrm{cov}(\alpha_t, y_t|Y_{t-1})$. The proof of the Kalman filter can be obtained by applying some basic results on the multivariate normal distribution or by applying linear prediction results; see, for example, Duncan and Horn (1972), Anderson and Moore (1979) and Harvey (1989).

The Kalman filter recursions for given κ are initialized by

$$a_1 = E(\alpha_1|Y_0) = E(\alpha_1) = a \qquad P_1 = \mathrm{var}(\alpha_1|Y_0) = \mathrm{var}(\alpha_1) = P \tag{4}$$

where a and P are the unconditional mean and variance matrix of the initial state vector, respectively. The diffuse case of $\kappa \to \infty$ is discussed when we consider the univariate form of the filter in Section 4.2.

2.3. *Smoothing*

Estimators of the state and disturbance vectors, conditional on the full set of observations $Y_n = \{y_1, \ldots, y_n\}$, are referred to as smoothed estimators and are

evaluated by backwards smoothing algorithms. The work of de Jong (1988), Kohn and Ansley (1989) and Koopman (1993) leads to the following basic smoothing recursions for model (1):

$$r_{t-1} = Z_t' F_t^{-1} v_t + L_t' T_t' r_t \qquad N_{t-1} = Z_t' F_t^{-1} Z_t + L_t' T_t' N_t T_t L_t \tag{5}$$

for $t = n, \ldots, 1$ with $L_t = I - K_t F_t^{-1} Z_t$. The backwards recursions (5) are initialized by $r_n = 0$ and $N_n = 0$. Storage of the Kalman filter output v_t, F_t^{-1} and K_t is required for $t = 1, \ldots, n$.

The output of recursions (5) can be used to construct the smoothed estimators of the disturbance vectors ε_t and η_t conditional on the full data set Y_n, i.e. $\hat{\varepsilon}_t = E(\varepsilon_t | Y_n)$ and $\hat{\eta}_t = E(\eta_t | Y_n)$, together with their variance matrices. These smoothed estimators are computed by

$$\begin{aligned} \hat{\varepsilon}_t = H_t F_t^{-1}(v_t - K_t' r_t) \qquad & \operatorname{var}(\hat{\varepsilon}_t) = H_t F_t^{-1}(F_t + K_t' N_t K_t) F_t^{-1} H_t \\ \hat{\eta}_t = Q_t R_t' r_t \qquad & \operatorname{var}(\hat{\eta}_t) = Q_t R_t' N_t R_t Q_t \end{aligned} \tag{6}$$

for $t = n, \ldots, 1$. The proofs and more general results for smoothed disturbances are given by Koopman (1993).

The smoothed state vector $\hat{\alpha}_t = E(\alpha_t | Y_n)$ and variance matrix $V_t = \operatorname{var}(\alpha_t | Y_n)$ also use (5) and can be evaluated by

$$\hat{\alpha}_t = a_t + P_t r_{t-1} \qquad V_t = P_t - P_t N_{t-1} P_t \tag{7}$$

for $t = n, \ldots, 1$. A substantial amount of additional memory space is required for the storage of a_t and P_t. Proofs of (5) and (7) are given by de Jong (1988) and Kohn and Ansley (1989). The state smoother (5) and (7) can also be obtained by re-formulating the classical Anderson and Moore (1979) fixed interval smoothing algorithm; see Koopman (1998).

A more efficient algorithm for calculating the smoothed estimator of the state vector only is given by

$$\hat{\alpha}_{t+1} = T_t \hat{\alpha}_t + R_t \hat{\eta}_t \qquad t = 1, \ldots, n \tag{8}$$

with $\hat{\alpha}_1 = a + P r_0$ and $\hat{\eta}_t$ given by (6). The forwards recursion (8) can be applied after the smoothing algorithm (5) has stored the vector r_t using the storage space of the Kalman filter, for $t = 1, \ldots, n$. The substantial storage space for the state smoother (7) is not required. Also, the recursion (8) is computationally more efficient than the first equation of (7) because the matrices T_t and R_t in (8) are usually sparse; see Koopman (1993) for a discussion.

3. UNIVARIATE APPROACH TO MULTIVARIATE CASE

Assuming first that variance matrix H_t is diagonal, write the observation and observation disturbance vectors as

$$y_t = \begin{pmatrix} y_{t,1} \\ \vdots \\ y_{t,p_t} \end{pmatrix} \qquad \varepsilon_t = \begin{pmatrix} \varepsilon_{t,1} \\ \vdots \\ \varepsilon_{t,p_t} \end{pmatrix}$$

with the observation system matrices

$$Z_t = \begin{pmatrix} Z_{t,1} \\ \vdots \\ Z_{t,p_t} \end{pmatrix} \qquad H_t = \begin{pmatrix} \sigma^2_{t,1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma^2_{t,p_t} \end{pmatrix}$$

where $y_{t,i}$, $\varepsilon_{t,i}$ and $\sigma^2_{t,i}$ are scalars and $Z_{t,i}$ is a $1 \times m$ row vector, for $i = 1, \ldots, p_t$. The observation equation for the univariate representation of the model is

$$y_{t,i} = Z_{t,i}\alpha_{t,i} + \varepsilon_{t,i} \qquad t = 1, \ldots, n; \qquad i = 1, \ldots, p_t \tag{9}$$

where $\alpha_{t,i} = \alpha_t$. The state equation corresponding to (9) is

$$\begin{aligned} \alpha_{t,i+1} &= \alpha_{t,i} \qquad i = 1, \ldots, p_t - 1 \\ \alpha_{t+1,1} &= T_t\alpha_{t,p_t} + R_t\eta_t \qquad t = 1, \ldots, n \end{aligned} \tag{10}$$

with initial state vector $\alpha_{1,1} = \alpha_1$ given by (2).

When H_t is not diagonal, the univariate representation of model (1) does not lead to an equivalent model because the correlations between the observation equations are lost. In this situation we can pursue two different approaches. First, we can put the disturbance vector ε_t into the state vector. For the observation equation of (1) define

$$\overline{\alpha}_t = \begin{pmatrix} \alpha_t \\ \varepsilon_t \end{pmatrix} \qquad \overline{Z}_t = (Z_t \quad I_{m_t})$$

and for the state equation define

$$\overline{\eta}_t = \begin{pmatrix} \eta_t \\ \varepsilon_t \end{pmatrix} \qquad \overline{T}_t = \begin{pmatrix} T_t & 0 \\ 0 & 0 \end{pmatrix} \qquad \overline{R}_t = \begin{pmatrix} R_t & 0 \\ 0 & I_{m_t} \end{pmatrix} \qquad \overline{Q}_t = \begin{pmatrix} Q_t & 0 \\ 0 & H_t \end{pmatrix}$$

leading to

$$y_t = \overline{Z}_t\overline{\alpha}_t \qquad \overline{\alpha}_{t+1} = \overline{T}_t\overline{\alpha}_t + \overline{R}_t\overline{\eta}_t \qquad \overline{\eta}_t \sim \mathrm{N}(0, \overline{Q}_t)$$

for $t = 1, \ldots, n$. We then proceed with the same strategy as for the case where H_t is diagonal by treating each element of the observation vector individually. The second approach is to transform the observations. In the case that H_t is not diagonal, we transform H_t by a singular value decomposition, i.e.

$$H^*_t = M_t H_t M'_t$$

where matrix H^*_t is diagonal. For example, Schur's decomposition let matrix M_t be orthogonal such that $M'_t M_t = I$; see Magnus and Neudecker (1988, Ch. 1,

Theorem 13). By transforming the observations, we obtain the observation equation

$$y_t^* = Z_t^* \alpha_t + \varepsilon_t^* \qquad \varepsilon_t^* \sim \mathrm{N}(0,\, H_t^*)$$

where $y_t^* = M_t y_t$, $Z_t^* = M_t Z_t$ and $\varepsilon_t^* = M_t \varepsilon_t$. The state vector is not affected by the transformion. Without further complications we can proceed with the univariate approach of filtering and smoothing, which we present in the next two sections.

These two approaches for correlated observation equations are complementary. The first method has the drawback that the state vector can potentially become large. The second method is illustrated in Section 7.2 where we further show that transforming the state vector as well can be convenient.

4. UNIVARIATE FILTERING

4.1. *The basic algorithm*

Define $a_{t,1} = E(\alpha_{t,1}|Y_{t-1})$ and $a_{t,i} = E(\alpha_{t,i}|Y_{t-1}, y_{t,1}, \ldots, y_{t,i-1})$ with $P_{t,1} = \mathrm{var}(\alpha_{t,1}|Y_{t-1})$ and $P_{t,i} = \mathrm{var}(\alpha_{t,i}|Y_{t-1}, y_{t,1}, \ldots, y_{t,i-1})$, for $i = 2, \ldots, p_t$. By treating the vector series $y_1, \ldots, y_n$ as the scalar series

$$y_{1,1}, \ldots, y_{1,p_t}, y_{2,1}, \ldots, y_{n,p_n}$$

the filtering equations where H_t is diagonal can be written as

$$a_{t,i+1} = a_{t,i} + K_{t,i} F_{t,i}^{-1} v_{t,i} \qquad P_{t,i+1} = P_{t,i} - K_{t,i} F_{t,i}^{-1} K_{t,i}' \tag{11}$$

where

$$v_{t,i} = y_{t,i} - Z_{t,i} a_{t,i} \qquad F_{t,i} = Z_{t,i} P_{t,i} Z_{t,i}' + \sigma_{t,i}^2 \qquad K_{t,i} = P_{t,i} Z_{t,i}' \tag{12}$$

for $i = 1, \ldots, p_t$ and $t = 1, \ldots, n$. This formulation has $v_{t,i}$ and $F_{t,i}$ as scalars and $K_{t,i}$ as a column vector. The transition from time t to time $t+1$ is achieved by the relations

$$a_{t+1,1} = T_t a_{t,p_t+1} \qquad P_{t+1,1} = T_t P_{t,p_t+1} T_t' + R_t Q_t R_t'. \tag{13}$$

These values $a_{t+1,1}$ and $P_{t+1,1}$ are the same as the values a_{t+1} and P_{t+1} given by the standard Kalman filter (3).

It is important to note that the elements of the innovation vector v_t of (3) are not the same as $v_{t,i}$ for $i = 1, \ldots, p_t$; only the first element of v_t is equal to $v_{t,1}$. The same applies to the diagonal elements of the variance matrix F_t and the variances $F_{t,i}$, for $i = 1, \ldots, p_t$; only the first diagonal element of F_t is equal to $F_{t,1}$. It is reasonable to assume that the full matrix F_t is not zero since this would indicate a model that had not been properly formulated. However, there are models for which $F_{t,i}$ can be zero, e.g. the case where y_t is a multinomial observation. This indicates that $y_{t,i}$ is linearly dependent on previous observations. Thus,

$$a_{t,i+1} = E(\alpha_{t,i+1}|Y_{t-1}, y_{t,1}, \ldots, y_{t,i}) = E(\alpha_{t,i+1}|Y_{t-1}, y_{t,1}, \ldots, y_{t,i-1}) = a_{t,i}$$

and similarly $P_{t,i+1} = P_{t,i}$. The contingency is therefore easily dealt with.

The main motivation of this univariate approach to filtering for multivariate state space models is computational efficiency. This approach avoids the inversion of matrix F_t and two matrix multiplications. Also, the implementation of the recursions is more straightforward. Table I shows that the percentage savings in the number multiplications for the univariate approach compared with the standard approach are considerable. The calculations concerning the transition (13) are not considered because matrix T_t is usually sparse with most elements equal to zero and unity.

4.2. *Diffuse filtering*

The filtering recursions (11)–(13) are valid for initial condition (2) with any fixed $\kappa > 0$. The diffuse case of $\kappa \to \infty$ requires some adjustments for a limited number of filtering steps until the dependence of $P_{t,i}$ on κ has vanished. The method of diffuse initialization is based on the treatment of Ansley and Kohn (1990) and Koopman (1997). The notation is similar to that adopted by Koopman (1997).

The definition $P = P_* + \kappa P_\infty$ in (2) implies that the matrix $P_{t,i}$, the vector $K_{t,i}$ and the scalar $F_{t,i}$ can be decomposed as

$$\begin{aligned} P_{t,i} &= P_{*,t,i} + \kappa P_{\infty,t,i} \\ K_{t,i} &= K_{*,t,i} + \kappa K_{\infty,t,i} \\ F_{t,i} &= F_{*,t,i} + \kappa F_{\infty,t,i} \end{aligned} \tag{14}$$

where

$$\begin{aligned} F_{*,t,i} &= Z_{t,i} P_{*,t,i} Z'_{t,i} + \sigma^2_{t,i} \qquad F_{\infty,t,i} = Z_{t,i} P_{\infty,t,i} Z'_{t,i} \\ K_{*,t,i} &= P_{*,t,i} Z'_{t,i} \qquad K_{\infty,t,i} = P_{\infty,t,i} Z'_{t,i} \end{aligned} . \tag{15}$$

TABLE I

PERCENTAGE COMPUTING SAVINGS FOR FILTERING

	p					
m	1	2	3	5	10	20
1	0	39	61	81	94	98
2	0	27	47	69	89	97
3	0	21	38	60	83	95
5	0	15	27	47	73	90
10	0	8	16	30	54	78
20	0	5	9	17	35	58

Percentages are calculated as $100(x - y)/x$ where x is the number of multiplications for the standard approach and y is the number of multiplications for the new univariate approach.

To obtain the diffuse filtering recursions, we expand $F_{t,i}^{-1}$ as a power series in κ^{-1} giving

$$F_{t,i}^{-1} = \kappa^{-1} F_{\infty,t,i}^{-1} - \kappa^{-2} F_{*,t,i} F_{\infty,t,i}^{-2} + \mathrm{O}(\kappa^{-3}) \qquad \text{for } F_{\infty,t,i} > 0.$$

This is easily obtained from the identity $F_{t,i}^{-1}(F_{*,t,i} + \kappa F_{\infty,t,i}) = 1$. From (11) the diffuse filtering recursions are therefore given by

$$\begin{aligned} a_{t,i+1} &= a_{t,i} + K_{\infty,t,i} F_{\infty,t,i}^{-1} v_{t,i} \\ P_{*,t,i+1} &= P_{*,t,i} + K_{\infty,t,i} K_{\infty,t,i}' F_{*,t,i} F_{\infty,t,i}^{-2} \\ &\quad - (K_{*,t,i} K_{\infty,t,i}' + K_{\infty,t,i} K_{*,t,i}') F_{\infty,t,i}^{-1} \\ P_{\infty,t,i+1} &= P_{\infty,t,i} - K_{\infty,t,i} K_{\infty,t,i}' F_{\infty,t,i}^{-1} \end{aligned} \tag{16}$$

for $i = 1, \ldots, p_t$. In the case where $F_{\infty,t,i} = 0$, the usual filtering equations apply, i.e.

$$\begin{aligned} a_{t,i+1} &= a_{t,i} + K_{*,t,i} F_{*,t,i}^{-1} v_{t,i} \\ P_{*,t,i+1} &= P_{*,t,i} - K_{*,t,i} K_{*,t,i}' F_{*,t,i}^{-1} \\ P_{\infty,t,i+1} &= P_{\infty,t,i} \end{aligned} \tag{17}$$

for $i = 1, \ldots, p_t$. For the transition from time t to $t+1$ we have

$$\begin{aligned} a_{t+1,1} &= T_t a_{t,p_t+1} \\ P_{*,t+1,1} &= T_t P_{*,t,p_t+1} T_t' + R_t Q_t R_t' \\ P_{\infty,t+1,1} &= T_t P_{\infty,t,p_t+1} T_t' \end{aligned} \tag{18}$$

for $t = 1, \ldots, n$.

Although it is not a restriction for a properly defined model, we require that

$$r(P_{\infty,t+1,1}) = r(P_{\infty,t,p_t+1}) \tag{19}$$

which implies that matrix T_t does not influence the rank of $P_{\infty,t,i}$. It can be shown that, when $F_{\infty,t,i} > 0$,

$$r(P_{\infty,t,i+1}) = r(P_{\infty,t,i}) - 1 \tag{20}$$

(see Ansley and Kohn, 1985, 1990; Koopman, 1997). The diffuse recursions (16)–(18) are continued until matrix $P_{\infty,t,i+1}$ becomes zero at t, $i = t^*$, i^*. From then on the usual Kalman filter is used with $P_{t,i+1} = P_{*,t,i+1}$, The univariate series

$$y_{1,1}, \ldots, y_{1,p_t}, y_{2,1}, \ldots, y_{t^*,i^*}$$

will be referred to as the initial series.

It can be shown that, when $F_{\infty,t,i} > 0$, the filtering recursion (16) for $P_{t,i}^{\dagger} = (P_{*,t,i}, P_{\infty,t,i})$ can be written compactly as

$$P^{\dagger}_{t,i+1} = L^{\dagger}_{t,i} P^{\dagger}_{t,i} \qquad \text{with } L^{\dagger}_{t,i} = \begin{pmatrix} L_{\infty,t,i} & L_{0,t,i} \\ 0 & L_{\infty,t,i} \end{pmatrix}, \quad i = 1, \ldots, p_t \tag{21}$$

where

$$\begin{aligned} L_{\infty,t,i} &= I - K_{\infty,t,i} Z_{t,i} F^{-1}_{\infty,t,i} \\ L_{0,t,i} &= (K_{\infty,t,i} F_{*,t,i} F^{-1}_{\infty,t,i} - K_{*,t,i}) Z_{t,i} F^{-1}_{\infty,t,i} \end{aligned} \tag{22}$$

(see Koopman and Durbin, 1999).

The diffuse filtering equations imply a limited number of additional multiplications compared with the usual Kalman filter. The computational implications are discussed in Koopman (1997) where it is argued that this method outperforms existing methods for univariate cases. This approach of diffuse multivariate filtering, which is similar to the device given by Ansley and Kohn (1990, Section 4), is simpler and computationally more efficient than the methods proposed by Ansley and Kohn (1985) and Koopman (1997), which require intricate Cholesky transformations on variance matrices such as P_t and F_t.

5. UNIVARIATE SMOOTHING

5.1. *The basic algorithm*

The basic smoothing recursions (5) for the model (1) can be reformulated for the univariate series

$$y_{1,1}, \ldots, y_{1,p_t}, y_{2,1}, \ldots, y_{n,p_n}$$

as

$$\begin{aligned} r_{t,i-1} &= Z'_{t,i} F^{-1}_{t,i} v_{t,i} + L'_{t,i} r_{t,i} & N_{t,i-1} &= Z'_{t,i} F^{-1}_{t,i} Z_{t,i} + L'_{t,i} N_{t,i} L_{t,i} \\ r_{t-1,p_t} &= T'_{t-1} r_{t,0} & N_{t-1,p_t} &= T'_{t-1} N_{t,0} T_{t-1} \end{aligned} \tag{23}$$

where $L_{t,i} = I - K_{t,i} Z_{t,i} F^{-1}_{t,i}$, for $i = p_t, \ldots, 1$ and $t = n, \ldots, 1$. The initializations are $r_{n,p_n} = 0$ and $N_{n,p_n} = 0$. The equations for r_{t-1,p_t} and N_{t-1,p_t} do not apply for $t = 1$. The values for $r_{t,0}$ and $N_{t,0}$ are the same as the values for the smoothing quantities r_{t-1} and N_{t-1} of (5), respectively.

The univariate smoothing approach avoids two matrix multiplications and the implementation is more straightforward. Table II presents the considerable percentage savings in the number of multiplications for the univariate approach compared with the standard multivariate approach. The computations involving the usually sparse transition matrix T_t are not considered.

5.2. *State and disturbance smoothing*

The state smoothing equations for the univariate approach provide the same results as Equations (7) since $a_t = a_{t,1}$, $P_t = P_{t,1}$, $r_{t-1} = r_{t,0}$ and $N_{t-1} = N_{t,0}$.

TABLE II

PERCENTAGE COMPUTING SAVINGS FOR SMOOTHING

	p					
m	1	2	3	5	10	20
1	0	27	43	60	77	87
2	0	22	36	53	72	84
3	0	19	32	48	68	81
5	0	14	25	40	60	76
10	0	9	16	28	47	65
20	0	5	10	18	33	51

Percentages are calculated as $100(x - y)/x$ where x is the number of multiplications for the standard approach and y is the number of multiplications for the new univariate approach.

Similar considerations apply for the smoothed disturbances $\hat{\eta}_t$ and $\text{var}(\hat{\eta}_t)$ in (6) and the state smoother (8). The smoothed estimators for the observation disturbances $\varepsilon_{t,i}$ of (9) follow directly from the univariate approach and are given by

$$\hat{\varepsilon}_{t,i} = \sigma^2_{t,i} F^{-1}_{t,i}(v_{t,i} - K'_{t,i} r_{t,i})$$

$$\text{var}(\hat{\varepsilon}_{t,i}) = \sigma^4_{t,i} F^{-2}_{t,i}(F_{t,i} + K'_{t,i} N_{t,i} K_{t,i}).$$

5.3. *Diffuse smoothing*

In this section we present the diffuse smoothing recursions for the initial series with indices

$$(t, i) = (t^*, i^*), (t^*, i^* - 1), \ldots, (t^*, 1), (t^* - 1, p_{t^*-1}), \ldots, (1, 1).$$

The treatment is based on Koopman and Durbin's (1999) results for the vector observation case.

To obtain smoothed estimators as $\kappa \to \infty$, we expand $r_{t,i}$ and $N_{t,i}$ of (23) in terms of reciprocals of κ in the same way as for $F^{-1}_{t,i}$, i.e.

$$\begin{aligned} r_{t,i} &= r^{(0)}_{t,i} + \kappa^{-1} r^{(1)}_{t,i} + \text{O}(\kappa^{-2}) \\ N_{t,i} &= N^{(0)}_{t,i} + \kappa^{-1} N^{(1)}_{t,i} + \kappa^{-2} N^{(2)}_{t,i} + \text{O}(\kappa^{-3}) \end{aligned} \tag{24}$$

with $r^{(0)}_{t^*,i^*} = r_{t^*,i^*}$, $r^{(1)}_{t^*,i^*} = 0$, $N^{(0)}_{t^*,i^*} = N_{t^*,i^*}$ and $N^{(1)}_{t^*,i^*} = N^{(2)}_{t^*,i^*} = 0$. We need three terms in the series for $N_{t,i}$ compared with two in the series for $r_{t,i}$ to allow for the contribution of terms in κ and κ^2 from the multiplications of $P_t = P_{*,t} + \kappa P_{\infty,t}$ required for state smoothing as given by (7). Note that r_{t^*,i^*} and N_{t^*,i^*} are obtained from (23) at $t, i = t^*, i^*$. By defining

$$r^{\dagger}_{t,i} = \begin{pmatrix} r^{(0)}_{t,i} \\ r^{(1)}_{t,i} \end{pmatrix} \qquad N^{\dagger}_{t,i} = \begin{pmatrix} N^{(0)}_{t,i} & N^{(1)}_{t,i} \\ N^{(1)}_{t,i} & N^{(2)}_{t,i} \end{pmatrix}$$

it can be shown using (23) that the diffuse basic smoothing equations, when $F_{\infty,t,i} > 0$, are given by

$$\begin{aligned} r^{\dagger}_{t,i-1} &= \begin{pmatrix} 0 \\ Z'_{t,i} F^{-1}_{\infty,t,i} v_{t,i} \end{pmatrix} + L^{\dagger\prime}_{t,i} r^{\dagger}_{t,i} \\ N^{\dagger}_{t,i-1} &= \begin{pmatrix} 0 & Z'_{t,i} F^{-1}_{\infty,t,i} Z_{t,i} \\ Z'_{t,i} F^{-1}_{\infty,t,i} Z_{t,i} & Z'_{t,i} F^{-2}_{\infty,t,i} Z_{t,i} F_{*,t,i} \end{pmatrix} + L^{\dagger\prime}_{t,i} N^{\dagger}_{t,i} L^{\dagger}_{t,i} \end{aligned} \tag{25}$$

where $L^{\dagger}_{t,i}$ is defined as in (21) for the initial series and with

$$\begin{aligned} r^{\dagger}_{t-1,p_t} &= \begin{pmatrix} T_{t-1} & 0 \\ 0 & T_{t-1} \end{pmatrix}' r^{\dagger}_{t,0} \\ N^{\dagger}_{t-1,p_t} &= \begin{pmatrix} T_{t-1} & 0 \\ 0 & T_{t-1} \end{pmatrix}' N^{\dagger}_{t,0} \begin{pmatrix} T_{t-1} & 0 \\ 0 & T_{t-1} \end{pmatrix} \end{aligned}$$

for $t = t^*, \ldots, 1$; see Koopman and Durbin (1999) for more computational details. It should be noted that the recursions (25) for $r^{\dagger}_{t,i-1}$ and $N^{\dagger}_{t,i-1}$ can be implemented in a computationally efficient way by taking account of symmetric structured matrices and duplicate matrices.

The diffuse state smoothing equations are given by

$$\hat{\alpha}_t = a_{t,1} + P^{\dagger}_{t,1} r^{\dagger}_{t,0} \qquad V_t = P_{*,t,1} - P^{\dagger}_{t,1} N^{\dagger}_{t,0} P^{\dagger\prime}_{t,1} \tag{26}$$

for $t = t^*, \ldots, 1$. The diffuse smoothed disturbances for the initial series are given by

$$\begin{aligned} \hat{\varepsilon}_{t,i} &= -\sigma^2_{t,i} F^{-1}_{\infty,t} K'_{\infty,t} r^{(0)}_{t,i} \qquad \mathrm{var}(\hat{\varepsilon}_{t,i}) = \sigma^4_{t,i} F^{-2}_{\infty,t} K'_{\infty,t} N^{(0)}_{t,i} K_{\infty,t} \\ \hat{\eta}_t &= Q_t R'_t r^{(0)}_{t,0} \qquad \mathrm{var}(\hat{\eta}_t) = Q_t R'_t N^{(0)}_{t,0} R_t Q_t \end{aligned} \tag{27}$$

where it should be noted that the smoothed disturbance equations (27) do not need the quantities $r^{(1)}_{t,i}$, $N^{(1)}_{t,i}$ and $N^{(2)}_{t,i}$, which simplifies the calculations considerably.

6. PARAMETER ESTIMATION

The system matrices Z_t, H_t, T_t, R_t and Q_t of model (1) may depend on unknown parameters that can be estimated by maximum likelihood. Let us denote the vector of these parameters by ψ. The output of the Kalman filter enables the evaluation of the log likelihood function via the prediction error decomposition for given ψ, and the score vector for ψ can be constructed using

the basic smoothing equations for a given vector ψ. Numerical optimization routines can be used to maximize the log likelihood function with respect to ψ.

The Gaussian log likelihood function for model (9) and (10) is given by

$$\log L = \text{constant} - 0.5 \sum_{t=1}^{n} \sum_{i=1}^{p_t} \log F_{t,i} + v_{t,i}^2 F_{t,i}^{-1} \tag{28}$$

where $v_{t,i}$ and $F_{t,i}$ are defined in Section 4.1. The log likelihood function (28) is obtained by treating the series of vector observations as a univariate series and applying the prediction error decomposition; see Harvey (1989, Section 3.4). The conventional method of log likelihood evaluation is based on the usual Kalman filter (3) and is given by

$$\log L = \text{constant} - 0.5 \sum_{t=1}^{n} \log|F_t| + v_t' F_t^{-1} v_t. \tag{29}$$

Equation (28) is computationally more efficient to compute than (29) because the univariate Kalman filter is more efficient and (28) avoids calculating the determinant of F_t.

The score vector for ψ can be obtained via the basic smoothing recursions (5) which may lead to dramatic computational efficiencies compared with numerical score evaluation; see Koopman and Shephard (1992). For example, let the ith element of ψ represent some unknown value of the system matrices R_t, for $t = 1, \ldots, n$. Its score value evaluated at $\psi = \psi^*$ is given by

$$\left.\frac{\partial \log L}{\partial \psi_i}\right|_{\psi=\psi^*} = \sum_{t=1}^{n} \text{tr} \frac{\partial R_t}{\partial \psi_i} Q_t R_t' (r_{t,0} r_{t,0}' - N_{t,0})$$

where $r_{t,0}$ and $N_{t,0}$ are defined in Section 5.1. Similar expressions exist for elements of ψ that are associated with system matrices H_t and Q_t. The equation for the score of a parameter that is associated with the system matrices Z_t and/or T_t is intricate and requires state smoothing. Koopman and Shephard (1992) argue that in this case it is computationally more efficient to compute the score numerically.

The log likelihood function for the diffuse case is given by

$$\log L = \text{constant} - 0.5 \sum_{t=1}^{t^*} \sum_{i=1}^{i^*} \log F_{\infty,t,i} - 0.5 \sum_{t=t^*}^{n} \sum_{i=i^*+1}^{p_t} \log F_{t,i} + v_{t,i}^2 F_{t,i}^{-1}.$$

In the example given earlier, the score for the diffuse case is given by

$$\left.\frac{\partial \log L}{\partial \psi_i}\right|_{\psi=\psi^*} = \sum_{t=1}^{n} \text{tr} \frac{\partial R_t}{\partial \psi_i} Q_t R_t' (r_{t,0}^{(0)} r_{t,0}^{(0)\prime} - N_{t,0}^{(0)}).$$

For all models used in the practical time series analysis, it is found that $F_{\infty,t,i}$ is independent of the unknown parameter vector ψ. The diffuse log likelihood and score functions are properly defined in such cases. Parameter estimation requires

many likelihood and score evaluations within the numerical optimization routine. It is fortunate that the auxiliary part of diffuse filtering, which consists of the equations for $F_{\infty,t,i}$, $K_{\infty,t,i}$ and $P_{\infty,t,i}$ does not depend on the system matrices H_t, R_t and Q_t. This follows immediately from a close examination of Equations (14)–(18). Therefore, the computations for $F_{\infty,t,i}$, $K_{\infty,t,i}$ and $P_{\infty,t,i}$ do not have to be repeated each time when a new likelihood evaluation is required for a new parameter vector ψ. This leads to considerable computational savings during the process of parameter estimation which cannot be achieved when one of the initialization strategies of de Jong (1991), Bell and Hillmer (1991) or Snyder and Saligari (1996) is adopted. By further examining the diffuse recursions and taking into account that most parameters associated with non-stationary or fixed unknown elements of the state vector do not affect the stationary part of the state vector, the computational efficiency also applies to parameters within ψ that are associated with T_t and Z_t.

7. APPLICATIONS

In this section we discuss three different applications in statistics and economics for which our results are particularly relevant. We do not give full numerical details; we only discuss the models and indicate why the univariate approach is superior to the standard approach.

7.1. *Multivariate time series models*

The state space model can be used for a variety of time series models such as the autoregressive moving average (ARMA) model, the unobserved components time series models and the dynamic regression model. The vector autoregressive (VAR) model and the multivariate structural time series model are further examples. State space representations of these models are discussed by Harvey (1989). The computational savings of these models are the same as for the general state space model and are given in Tables I and II. The computations involving the transition matrix T_t are not considered because of the sparse nature of this matrix for most models.

7.2. *Vector splines*

The generalization of smoothing splines (see Hastie and Tibshirani, 1990) to the multivariate case are considered by Fessler (1991) and Yee and Wild (1996). The vector spline model is given by

$$y_i = \theta(x_i) + \varepsilon_i \qquad E(\varepsilon_i) = 0 \qquad \text{var}(\varepsilon_i) = \Sigma_i \qquad i = 1, \ldots, n$$

where y_i is a $p \times 1$ vector response at scalar x_i, $\theta(\cdot)$ is an arbitrary smooth vector function and error ε_i is mutually uncorrelated. The variance matrix Σ_i is assumed to be known and is usually constant for varying i. The standard method

of estimating the smooth vector function is by minimizing the generalized least squares criterion

$$\sum_{i=1}^{n}\{y_i-\theta(x_i)\}'\Sigma_i^{-1}\{y_i-\theta(x_i)\}+\sum_{j=1}^{p}\lambda_j\int\theta_j''(x)^2\,dx$$

where the non-negative smoothing parameter λ_j determines the smoothness of the jth smooth function $\theta_j(\cdot)$ of vector $\theta(\cdot)$ for $j=1,\ldots,p$. Note that $x_{i+1}>x_i$ for $i=1,\ldots,n-1$ and $\theta_j''(x)$ denotes the second derivative of $\theta_j(x)$ with respect to x. In the same way as Wecker and Ansley (1983) put smoothing splines into state space form, vector splines can be represented as

$$y_i=\mu_i+\varepsilon_i$$
$$\mu_{i+1}=\mu_i+\delta_i\beta_i+\eta_i \qquad \text{var}(\eta_i)=(\delta_i^3/3)\Lambda$$
$$\beta_{i+1}=\beta_i+\zeta_i \qquad \text{var}(\zeta_i)+\delta_i\Lambda \qquad \text{cov}(\eta_i,\zeta_i)=(\delta_i^2/2)\Lambda$$

with vector $\mu_i=\theta(x_i)$, scalar $\delta_i=x_{i+1}-x_i$ and diagonal matrix $\Lambda=\text{diag}(\lambda_1,\ldots,\lambda_p)$. This model is equivalent to the continuous-time representation of the mulitivariate local linear trend model with no disturbance vector for the level equation; see Harvey (1989, Ch. 8). In the case of $\Sigma_i=\Sigma$ and diagonalization $M\Sigma M'=D$ where matrix M is orthogonal and matrix D is diagonal, we obtain the transformed model

$$y_i^*=\mu_i^*+\varepsilon_i^*$$
$$\mu_{i+1}^*=\mu_i^*+\delta_i\beta_i^*+\eta_i^* \qquad \text{var}(\eta_i)=(\delta_i^3/3)Q$$
$$\beta_{i+1}^*=\beta_i^*+\xi_i^* \qquad \text{var}(\zeta_i)=\delta_iQ \qquad \text{cov}(\eta_i,\zeta_i)=(\delta_i^2/2)Q$$

with $y_i^*=My_i$ and $\text{var}(\varepsilon_i^*)=D$. Furthermore, we have $\mu_i^*=M\mu_i$, $\beta_i^*=M\beta_i$ and $Q=M\Lambda M'$. The Kalman filter smoother algorithm provides the fitted smoothing spline. The untransformed model and the transformed model can both be handled by the univariate strategy of filtering and smoothing. The advantage of the transformed model is that ε_i^* can be excluded from the state vector, which is not possible for the untransformed model because $\text{var}(\varepsilon_i)=\Sigma_i$ is not necessarily diagonal; see the discussion in Section 3.

The percentage computational saving of the univariate approach for spline smoothing depends on the size p. The state vector dimension for the transformed model is $m=2p$ so that the percentage saving in computing for filtering is 30 if $p=5$ and 35 if $p=10$; see Table I. The percentages for smoothing are 28 and 33, respectively; see Table II.

7.3. *Modelling bid–ask spreads*

Competitive dealership markets, such as the London Stock Exchange and the Chicago Mercantile Exchange, have typically several dealers negotiating and completing multiple trades at the same time. Different market prices of the same

equity float within the market at the same period of, say, a minute. The sequential order of market prices in the same period is unknown. Moreover, the number of trades varies for different periods. Therefore the standard approach of disentangling the bid–ask spread from trade prices using the autocovariance structure of differenced market prices is not possible; see Huang and Stoll (1997) for an overview of the standard approach. Koopman and Lai (1998) offer an alternative approach by modelling the price data using a simple state space framework which deals with the specific features of competitive dealership markets. They apply their model using equity prices of Shell, Glaxo and British Telecom traded at the London Stock Exchange.

The basic specification of the model used by Koopman and Lai (1998) is

$$\begin{aligned} y_{t,i} &= \mu_t + d_{t,i}\alpha + \varepsilon_{t,i} \qquad \varepsilon_{t,i} \sim \mathrm{N}(0, \sigma_\varepsilon^2) \qquad i = 1, \ldots, p_t \\ \mu_{t+1} &= \mu_t + \eta_t \qquad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2) \qquad t = 1, \ldots, n \end{aligned} \tag{30}$$

where $y_{t,i}$ is a univariate series of equity prices and $d_{t,i}$ is zero or unity depending on whether the ith trade at time t is a buy or a sell. The spread is the constant α and the disturbances $\varepsilon_{t,i}$ are mutually independent and uncorrelated with the disturbances η_t. The number of trades within time period t, p_t, typically ranges from 0 to 100. The time index t is usually measured in seconds, minutes or quarters of hours. For example, the London Stock Exchange can provide trade information each minute. Various generalizations may be applied to this model. For example, the spread α can be a random walk with regression spline effects for time and trade size and the underlying 'true' price μ_t may be corrected for adverse selection effects; see Koopman and Lai (1998).

The univariate strategy of Kalman filtering and smoothing will dramatically decrease the number of computations for model (30) compared with the standard approach for this model. Tables I and II give the percentage savings for values of p_t up to 20 (and with $m = 1$ as for this model) but in this application p_t repeatedly takes values of 70 and more leading to even more dramatic savings, such as 99.96%. The size of n is typically in thousands so the computational savings are important in such applications.

8. CONCLUSION

In this paper we have considered filtering, smoothing and log likelihood estimation for multivariate linear state space models. We show that by bringing in elements of the observational vectors one by one instead of together as vectors, considerable, and in some cases spectacular, computational savings can be made. The exact treatment of diffuse priors in multivariate cases is simplified considerably by this univariate approach.

ACKNOWLEDGEMENTS

We would like to thank an anonymous referee for helpful comments. S. J. Koopman was a Research Fellow of the Royal Netherlands Academy of Arts and Sciences during most part of this project. Its financial support is gratefully acknowledged.

REFERENCES

ANDERSON, B. D. O. and MOORE, J. B. (1979). *Optimal Filtering*. Englewood Cliffs, NJ: Prentice Hall.

ANSLEY, C. F. and KOHN, R. (1985). Estimation, filtering and smoothing in state space models with incompletely specified initial conditions. *Ann. Stat*. 13, 1286–316.

—— and —— (1990). Filtering and smoothing in state space models with partially diffuse initial conditions. *J. Time Ser. Anal*. 11, 275–93.

BELL, W. and HILLMER, S. (1991). Initializing the Kalman filter for nonstationary time series models. *J. Time Ser. Anal*. 12, 283–300.

DUNCAN, D. B. and HORN, S. D. (1972). Linear dynamic regression from the viewpoint of regression analysis. *J. Am. Stat. Assoc*. 67, 815–21.

FAHRMEIR, L. and TUTZ, G. (1996). *Mutlivariate Statistical Modelling Based on Generalized Linear Models*. New York: Springer.

FESSLER, J. A. (1991). Nonparametric fixed-interval smoothing with vector splines. *IEEE Trans. Signal Process* 39, 852–59.

HARVEY, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

—— and DURBIN, J. (1986). The effects of seat belt legislation on British road casualties: a case study in structural time series modelling (with discussion). *J. R. Stat. Soc. A* 149, 187–227.

HASTIE, T. and TIBSHIRANI, R. (1990). *Generalized Additive Models*. London: Chapman and Hall.

HUANG, R. and STOLL, H. (1997). The components of bid–ask spread: a general approach. *Rev. Financ. Stud*. 10, 995–1034.

DE JONG, P. (1988). A cross validation filter for time series models. *Biometrika* 75, 594–600.

—— (1991). The diffuse Kalman filter. *Ann. Stat*. 19, 1073–83.

KOHN, R. and 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. (1993). Disturbance smoother for state space models. *Biometrika* 80, 117–26.

—— (1997). Exact initial Kalman filtering and smoothing for nonstationary time series models. *J. Am. Stat. Assoc*. 92, 1630–38.

—— (1998). Kalman filtering and smoothing. In *The Encyclopedia of Biostatistics* (eds P. Armitage and T. Colton). Chichester: Wiley.

—— and DURBIN, J. (1999). Diffuse state smoothing for state space models. Working paper.

—— and LAI, H. N. (1998). Modelling bid–ask spreads in competitive dealership markets. Working paper.

—— and SHEPHARD, N. (1992). Exact score for time series model in state space form. *Biometrika*, 79, 823–26.

MAGNUS, J. R. and NEUDECKER, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Chichester: Wiley.

SNYDER, R. D. and SALIGARI, G. R. (1996). Initialization of the Kalman filter with partially diffuse initial conditions. *J. Time Ser. Anal*. 17, 409–24.

WECKER, W. E. and ANSLEY, C. F. (1983). The signal extraction approach to nonlinear regression and spline smoothing. *J. Am. Stat. Assoc.* 78, 81–89.

YEE, T. W. and WILD, C. J. (1996). Vector generalized additive models. *J. R. Stat. Soc. B* 58, 481–93.