The Annals of Statistics
1985, Vol. 13, No. 4, 1286–1316

ESTIMATION, FILTERING, AND SMOOTHING IN STATE SPACE MODELS WITH INCOMPLETELY SPECIFIED INITIAL CONDITIONS

BY CRAIG F. ANSLEY AND ROBERT KOHN

University of Chicago

The likelihood is defined for a state space model with incompletely specified initial conditions by transforming the data to eliminate the dependence on the unspecified conditions. This approach is extended to obtain estimates of the state vectors and predictors and interpolators for missing observations. It is then shown that this method is equivalent to placing a diffuse prior distribution on the unspecified part of the initial state vector, and modified versions of the Kalman filter and smoothing algorithms are derived to give exact numerical procedures for diffuse initial conditions. The results are extended to continuous time models, including smoothing splines and continuous time autoregressive processes.

1. Introduction. In this paper we consider observations generated by a multivariate Gaussian state space model observed at discrete points in time. The initial state vector $x(0)$ will be of the form

$$x(0) = \Phi\eta + \zeta,$$

where ζ has a well-defined Gaussian distribution, while the distribution of η is unspecified. Although the likelihood of the observations cannot be defined in the usual sense because the distribution of η is unspecified, we show how to define a likelihood by taking a singular transformation of the data that eliminates the dependence on η. This likelihood will be invariant under a large class of transformations.

We also show how to use the observed data to estimate the unobserved state vector, how to predict future observations and interpolate missing observations, and how to obtain the mean-squared errors of all these estimates. Because the distribution of η is unspecified, we cannot simply obtain the above estimates by taking conditional expectations based on the data. Instead, we extend the transformation approach, mentioned in relation to the likelihood, to define estimates of the unobserved state vectors and the unobserved values of the dependent variables.

Although the definition of the likelihood as described above is conceptually satisfactory, the resulting expression for the likelihood will not usually be in a suitable form for efficient computation. This is because transforming the data will destroy the state space structure of the observations and so we will no longer be able to use the Kalman filter to efficiently compute the likelihood as in

Received December 1983; revised December 1984 and May 1985.
AMS 1980 *subject classifications.* Primary 62M15; Secondary 62M20, 60G35
Key words and phrases. State space, Kalman filter, smoothing, nonstationarity, exact likelihood, missing data, continuous time process.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics.* www.jstor.org

Schweppe (1965) and Jones (1980). Similar remarks apply to the estimate of the state vector, predictors, and interpolators mentioned above.

To obtain efficient algorithms we will take

$$\eta \sim N(0, kI) \tag{1.1}$$

and let $k \to \infty$, making η diffuse and hence the initial state vector $x(0)$ partially diffuse. We show that the likelihood as defined above is equivalent to assuming (1.1) and considering a suitably normed limit of the likelihood of the (untransformed) observations. We generalize Schweppe's (1965) approach for computing the likelihood of a Gaussian state space model by deriving a modified Kalman filter algorithm that allows the initial state vector to be partially diffuse. Although for fixed k, we could use the ordinary Kalman filter to compute the likelihood, what we must do is let $k \to \infty$, and this is not possible using the ordinary Kalman filter *because filtering and limiting operations are not interchangeable*. The modified filter shows explicitly how the conditional state covariance matrices and the innovation variances depend on k.

We will also show that the estimates described above of the state vector and the unobserved values of the dependent variable are limits (as $k \to \infty$) of conditional expectations based on the observed data under (1.1). Efficient computation is again done by modified smoothing algorithms, which allow the state vector to be partially diffuse. In particular, we obtain modified versions of the fixed interval and fixed point smoothing algorithms. As before, we need modified smoothing algorithms because smoothing and limiting operations are not interchangeable with the usual algorithms.

Although most of our results are derived for a discrete time state space model, we show how to extend them to a continuous time state space model observed at discrete points in time.

We have applied the results in this paper to compute the likelihood and predict and interpolate data in a nonstationary ARIMA model (Kohn and Ansley, 1984a) and to compute optimal smoothing splines (Kohn and Ansley, 1984b).

Other ways of handling a partially diffuse prior have been suggested in the literature and are discussed at the end of Section 6. They seem inferior to our approach.

This paper is structured as follows. Section 2 defines our model and states the assumptions. In Section 3 we define the likelihood and in Section 4 we show how to define estimates of the state vector and predict and interpolate missing observations. In Section 5 we assume, in addition, that (1.1) holds and show that the normalized likelihood now obtained is equivalent to that in Section 3. In Section 6 we state the modified Kalman filter and in Section 7 we apply it to efficiently compute the likelihood. In Section 8 we discuss smoothing algorithms and in Section 9 we outline how our results can be extended to a continuous time state space model observed at discrete points in time.

Because the proofs of the results in Sections 6–8 are rather long and require a number of preliminary results, we have placed both the preliminary results and the proofs in Section 10 at the end of the paper.

2. Model and assumptions. We consider the $p \times 1$ vector stochastic process $(y(t), t \geq 1)$ generated by the state space model:

$$y(t) = H(t;\theta)x(t) + v(t), \qquad x(t+1) = F(t;\theta)x(t) + u(t). \tag{2.1}$$

ASSUMPTION 2.1. (i) $(x(t), t \geq 1)$ is a sequence of $q \times 1$ state vectors.

(ii) For each t: (a) $v(t)$ and $u(t)$ are, respectively, $p \times 1$ and $q \times 1$ Gaussian random disturbances, both having zero mean and with variance–covariance matrices $Q(t;\theta)$ and $R(t;\theta)$, respectively; (b) $v(t)$ and $u(t)$ are independent of $x(s)$ for all $s \leq t$; (c) $v(t)$ and $u(s)$ are independent for all s and t; (d) $v(t)$ $(t \geq 1)$ is an independent sequence as is $u(t)$ $(t \geq 1)$.

(iii) The matrices $H(t;\theta)$ and $F(t;\theta)$ are $p \times q$ and $q \times q$, respectively for $t \geq 1$.

ASSUMPTION 2.2 (Initial Conditions).

$$x(0) = \Phi(\theta)\eta + \zeta, \tag{2.2}$$

where

(i) $\Phi(\theta)$ is a $q \times D$ $(D \leq q)$ matrix.

(ii) $\zeta \sim N(0, \Sigma(\theta))$ and η is a $D \times 1$ vector having an unspecified distribution. Both ζ and η are independent of $u(0)$ and $(v(t), u(t); t \geq 1)$.

ASSUMPTION 2.3. θ is a parameter vector belonging to a subset Θ of a finite dimensional Euclidean space.

ASSUMPTION 2.4. We observe $y(t_1), \dots, y(t_n)$ with $1 = t_1 < t_2 < \cdots < t_n = N$.

Now let $\nu(0) = \zeta$ and for $t \geq 1$ define

$$\begin{aligned} A(t) &= \left(H(t) \prod_{j=0}^{t-1} F(j) \right) \Phi, \\ \nu(t) &= F(t-1)\nu(t-1) + u(t-1), \qquad \omega(t) = H(t)\nu(t) + v(t). \end{aligned} \tag{2.3}$$

Then

$$y(t) = A(t)\eta + \omega(t), \qquad t \geq 1. \tag{2.4}$$

Define the $n \times 1$ vectors $y = (y(t_1)', \dots, y(t_n)')'$ and $\omega = (\omega(t_1)', \dots, \omega(t_n)')'$, and let A be the $np \times D$ matrix having rows $(t-1)p+1, \dots, tp$ given by $A(t)$. Then

$$y = A\eta + \omega. \tag{2.5}$$

We assume that:

ASSUMPTION 2.5. (i) A is independent of θ.

(ii) The variance–covariance matrix of ω is nonsingular for all $\theta \in \Theta$.

(2.4) and Assumption 2.5(i) together imply that the dependence of $y(t)$ on η does not involve θ, although Φ, $H(j)$, and $F(j)$ $(j = 0, \dots, t-1)$ may depend on θ.

We can often check Assumptions 2.5(i) and (ii) from first principles. A sufficient condition for Assumption 2.5(ii) to hold is that $Q(t;\theta)$ is nonsingular for $t \geq 1$ and all $\theta \in \Theta$. The following lemma gives sufficient conditions for Assumption 2.5(i) to hold.

LEMMA 2.1. *A sufficient condition for Assumption* 2.5(i) *to hold is that* Φ, $F(t)$, *and* $H(t)$ $(t \geq 0)$ *are of the form*

$$\Phi = \begin{bmatrix} \Phi_1 \\ \cdots \\ 0 \end{bmatrix}, \qquad F(t) = \begin{bmatrix} F_{11}(t) & F_{12}(t) \\ 0 & F_{22}(t) \end{bmatrix}, \qquad H(t) = (H_1(t), H_2(t)),$$

where Φ_1 *and* $F_{11}(t)$ *are* $D \times D$, $H_1(t)$ *is* $p \times D$, *and* Φ_1, $F_{11}(t)$, *and* $H_1(t)$ $(t \geq 0)$ *are independent of* θ.

The proof is immediate. □

To illustrate our results we apply them throughout the paper to a nonstationary quarterly ARIMA model. A complete treatment of ARIMA models by the methods of this paper is given in Kohn and Ansley (1984a), with algorithmic details set out in Ansley and Kohn (1985).

EXAMPLE 2.1. Consider the ARIMA model

$$y(t) = y(t-4) + e(t) + \alpha e(t-1) \tag{2.6}$$

with $e(t)$ a sequence of independent $N(0, \sigma^2)$ random variables.

As in Kohn and Ansley (1984a) we can write (2.6) in state space form as

$$y(t) = Hx(t), \qquad x(t+1) = Fx(t) + g(\theta)e(t+1),$$

where $\theta = (\alpha, \sigma^2)$, $x(t)$ is a 4×1 state vector with elements $x_1(t) = y(t)$, $x_2(t) = y(t-3) + \alpha e(t)$, $x_3(t) = y(t-2)$, and $x_4(t) = y(t-1)$; $H = (1,0,0,0)$, $g' = (1, \alpha, 0, 0)$, and F is a 4×4 matrix with $F_{4,1} = 1$, $F_{j,j+1} = 1$ $(j = 1, \ldots, 3)$ and the rest of the elements of F are zero.

We define $\eta = (y(-3), \ldots, y(0))'$ and because $y(t)$ is nonstationary the distribution of η is unspecified. We can write $x(0)$ as in (2.2) with $\zeta = (0, \alpha e(0), 0, 0)'$ and

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

It is clear that for any sequence of observations $y(t_1), \ldots, y(t_n)$ generated by (2.6), Assumption 2.5 holds. For example, if we observe

$$y = (y(1), y(4), y(5), y(6), y(8), y(9), y(10), y(12))', \tag{2.7}$$

then $n = 8$ and

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{2.8}$$

3. Defining the likelihood. Because the density of η is unspecified, so is the density of y and hence the likelihood of the observations cannot be defined in the usual way.

To define a likelihood we transform y to eliminate dependence on η as follows. Let D' $(\le D)$ be the rank of A, and let J be an $n \times n$ matrix independent of θ so that $\det J = 1$ and JA has exactly D' nonzero rows. Because A is independent of θ, we can always construct such a matrix J. Let J_1 consist of those D' rows of J corresponding to the nonzero rows of JA and let J_2 consist of the other rows of J, so that $J_2 A = 0$. Put $w_1 = J_1 y$ and $w_2 = J_2 y$. Then

$$w_1 = J_1 A\eta + J_1\omega, \qquad w_2 = J_2\omega,$$

and we will define the likelihood as the density of w_2. This likelihood is well defined because it does not depend on η. We show in Corollary 5.1 that this definition of the likelihood is invariant to the choice of J.

One way of obtaining J is to note that by Lemma 10.1 of Section 10 we can factorize AA' as $L\Delta L'$ where L is a lower triangular matrix having 1s on the diagonal and Δ is a diagonal matrix. Taking $J = L^{-1}$ gives us a suitable J.

We motivate our definition of the likelihood by considering Example 2.1.

EXAMPLE 2.1 (ctd). Suppose we observe $y(1), y(2), \dots, y(n)$ $(n > 4)$. Let $w(t) = y(t)$ $(t = 1, \dots, 4)$ and $w(t) = y(t) - y(t-4)$ $(t = 5, \dots, n)$, and put $w_1 = (w(1), \dots, w(4))'$ and $w_2 = (w(5), \dots, w(n))'$. Then $w_1 = J_1 y$ and $w_2 = J_2 y$, where $J_1 = [I_4, 0]$ and J_2 is the $(n-4) \times n$ matrix

$$J_2 = \begin{bmatrix} -1 & 0 & 0 & 0 & 1 & & & & 0 \\ & \ddots & & & & \ddots & & & \\ 0 & & & & -1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Let J be the $n \times n$ matrix $J = [J_1', J_2']'$; $\det J = 1$.

We define the likelihood as the density of the $n - 4$ differenced observations $w_2 = J_2 y$. Thus, our definition corresponds to that usually adopted for ARIMA models when there are no missing observations. See, for example, Box and Jenkins [(1976), Chapter 6].

If the observation vector y is given by (2.7), so that some observations are missing, we can no longer difference the data to define the likelihood but we can proceed as above with $D' = 3$ because rank $A = 3$ in (2.8).

4. Interpolation, extrapolation, and state estimation. In this section we show how to define estimates of unobserved $y(t)$ and $x(t)$ based on the observed data y, and also obtain the mean-squared estimation errors. Because the density of η is unspecified, so are the densities of y, $y(t)$, and $x(t)$ and we cannot obtain our estimates by simply taking conditional expectations based on the observations y. Instead we proceed as follows.

4.1. Interpolation and extrapolation. Let $Y(t)$ be a $p' \times 1$ subvector of $y(t)$. From (2.4) we can write

$$Y(t) = \tilde{A}(t)\eta + \tilde{\omega}(t)$$

with $\tilde{\omega}(t)$ the corresponding $p' \times 1$ subvector of $\omega(t)$ and $\tilde{A}(t)$ the $p' \times D$ matrix consisting of the corresponding p' rows of $A(t)$.

Suppose $Y(t)$ is unobserved and we want to estimate it using the data y and obtain the mean-squared estimation error. If $t > N$ we are predicting and if $1 \le t \le N$ we are interpolating.

Suppose the rows of $\tilde{A}(t)$ belong to the row space of A. Then there is a $p' \times n$ matrix $\tilde{a}(t)$ so that

$$\tilde{A}(t) = \tilde{a}(t)A.$$

Hence $Y(t) - \tilde{a}(t)y = \tilde{\omega}(t) - \tilde{a}(t)\omega = \tilde{w}(t)$ say. Then $\tilde{w}(t)$ and w_2 (defined in Section 3) have a proper joint distribution because neither depends on η. We define the estimate of $Y(t)$ given y as

$$\hat{Y}(t|N) = \tilde{a}(t)y + E(\tilde{w}(t)|w_2). \tag{4.1}$$

The mean-squared prediction error is

$$\begin{aligned} \hat{S}_Y(t|N) &= \operatorname{Var}(Y(t) - \hat{Y}(t|N)) \\ &= \operatorname{Var}(\tilde{w}(t)|w_2). \end{aligned} \tag{4.2}$$

If the rows of $\tilde{A}(t)$ do not belong to the row space of A, then we cannot estimate $Y(t)$ from y using our approach because we cannot find an $\tilde{a}(t)$ so that $Y(t) - \tilde{a}(t)y$ is independent of η. To obtain an estimate of $Y(t)$, and also its mean-squared estimation error, we would need to impose some additional (nondiffuse) distributional assumptions on η.

Although (4.1) and (4.2) give us formulae to estimate $Y(t)$ and obtain its mean-squared error, the direct evaluation of $E(\tilde{w}(t)|w_2)$ and $\operatorname{Var}(\tilde{w}(t)|w_2)$ will be computationally inefficient. This is because in transforming y to obtain w_2, we destroy the state space structure of the observations and so we can no longer use the efficient filtering and smoothing algorithms that are available for state space models. We overcome this difficulty with modified filtering and smoothing procedures in Sections 6–8.

To illustrate our definitions we go back to Example 2.1.

EXAMPLE 2.1 (ctd). Let y be given by (2.7) and A by (2.8). We can predict, for example, $y(13)$ and interpolate $y(2)$ because $\tilde{A}(13) = (1,0,0,0)$ and $\tilde{A}(2) = (0,1,0,0)$ lie in the row space of A. We cannot predict $y(15)$ or interpolate $y(3)$

because $\tilde{A}(15) = \tilde{A}(3) = (0, 0, 1, 0)$ are not in the row space of A as we have no third quarter observations.

If, in addition, $y(15)$ is observed, then the matrix A will have full rank and we can predict and interpolate all missing observations.

4.2. Estimating the state vector. Now let $\nu(t)$ be defined as in (2.3), and define $B(0) = \Phi$,

$$B(t) = \prod_{j=0}^{t=1} F(j)\Phi, \qquad t > 0. \tag{4.3}$$

Then

$$x(t) = B(t)\eta + \nu(t).$$

Now let $X(t)$ be a $q' \times 1$ subvector of $x(t)$. Then we can write

$$X(t) = \tilde{B}(t)\eta + \tilde{\nu}(t),$$

where $\tilde{\nu}(t)$ is the corresponding $q' \times 1$ subvector of $\nu(t)$ and $\tilde{B}(t)$ is the $q' \times D$ matrix consisting of the corresponding q' rows of $B(t)$. We can define the estimate of the vector $X(t)$ based on all the data and the mean-squared error of the estimate similarly to the way we defined $\hat{Y}(t|n)$ and $\hat{S}_Y(t|n)$ in Section 4.1.

If the rows of $\tilde{B}(t)$ belong to the row space of A, then there exists a $q' \times n$ matrix $\tilde{b}(t)$ such that $\tilde{B}(t) = \tilde{b}(t)A$ and so $X(t) - \tilde{b}(t)y = \tilde{\nu}(t) - \tilde{b}(t)\omega$. Analogously to (4.1) and (4.2), we define the estimate of $X(t)$ given y as

$$\hat{X}(t|n) = \tilde{b}(t)y + E(\tilde{\nu}(t) - \tilde{b}(t)\omega|w_2) \tag{4.4}$$

with mean squared error

$$\hat{S}_X(t|n) = \mathrm{Var}(\tilde{\nu}(t) - \tilde{b}(t)\omega|w_2).$$

If the rows of $\tilde{B}(t)$ do not belong to the row space of A, then we cannot estimate $X(t)$ as in (4.4), although we may be able to estimate some subvector of it.

5. Model with partially diffuse initial conditions. We now consider the state space model of Section 2 and, in addition, assume that

ASSUMPTION 5.1. $\eta \sim N(0, kI)$ $(k > 0)$ and satisfies Assumption 2.2.

To indicate our lack of knowledge about η, we will let $k \to \infty$ making η diffuse and $x(0)$ partially diffuse. For fixed $k > 0$ the likelihood of y is well defined, and we show that in the limit as $k \to \infty$, this likelihood (suitably normalized to avoid degeneracy) is equivalent to that given in Section 3.

THEOREM 5.1. *Let $f(w_2; \theta)$ be the density of w_2, where w_2 is defined in Section* 3 *and let $f(y; \theta; k)$ be the density of y for fixed $k > 0$. Then*

$$f(w_2; \theta) = c \lim_{k \to \infty} k^{D'/2} f(y; \theta; k), \tag{5.1}$$

where c is a constant independent of θ.

PROOF. Let $f(w_1; \theta; k)$ be the density of w_1, where w_1 is defined in Section 3 and let $f(w_2|w_1; \theta; k)$ be the conditional density of w_2 given w_1. Then it is not difficult to show that

$$\lim_{k\to\infty} k^{D'/2} f(w_1; \theta; k) = (2\pi)^{-D'/2}(\det J_1 AA'J_1')^{-1/2}$$

and

$$\lim_{k\to\infty} f(w_2|w_1; \theta; k) = f(w_2; \theta).$$

Now

$$k^{D'/2} f(y; \theta; k) = k^{D'/2} f(w_1; \theta; k) f(w_2|w_1; \theta; k)$$

and (5.1) follows because $J_1 AA'J_1'$ is independent of θ. □

COROLLARY 5.1. *Up to a constant independent of θ, the likelihood $f(w_2; \theta)$ as defined in Section 3 is invariant to the transformation J.*

PROOF. This follows directly from (5.1). □

Let $Y(t)$ and $X(t)$ be the subvectors of $y(t)$ and $x(t)$, respectively, defined in Section 4. For fixed $k > 0$, let $Y(t|N; k) = E(Y(t)|y)$ and $S_Y(t|N; k) = \operatorname{Var}(y(t)|y)$, and define $X(t|N; k)$ and $S_X(t|N; k)$ similarly with respect to $X(t)$. We show that if the estimate $\hat{Y}(t|N)$ defined in Section 4.1 exists, then

$$\hat{Y}(t|N) = \lim_{k\to\infty} Y(t|N; k), \qquad \hat{S}_Y(t|N) = \lim_{k\to\infty} S_Y(t|N; k). \tag{5.2}$$

Similarly, if the estimate $\hat{X}(t|N)$ defined in Section 4.2 exists, then

$$\hat{X}(t|N) = \lim_{k\to\infty} X(t|N; k), \qquad \hat{S}_X(t|N) = \lim_{k\to\infty} S_X(t|N; k). \tag{5.3}$$

THEOREM 5.2. (i) *Suppose the rows of $\tilde{A}(t)$ lie in the column space of A. Let $f(\tilde{w}(t)|w_2; \theta)$ be the conditional density of $\tilde{w}(t)$ given w_2, and $f(Y(t)|y; \theta; k)$ the conditional density of $Y(t)$ given y. $\tilde{A}(t)$ and $\tilde{w}(t)$ are defined in Section 4.1. Then*

$$f(\tilde{w}(t)|w_2; \theta) = \lim_{k\to\infty} f(Y(t)|y; \theta; k) \tag{5.4}$$

and (5.2) holds.

(ii) *Suppose the rows of $\tilde{B}(t)$ lie in the row space of A. Let $f(\tilde{\nu}(t) - \tilde{b}(t)\omega|w_2)$ be the conditional density of $\tilde{\nu}(t) - \tilde{b}(t)\omega$ given w_2, and let $f(X(t)|y; \theta; k)$ be the conditional density of $X(t)$ given y. $\tilde{B}(t)$, $\tilde{\nu}(t)$, and $\tilde{b}(t)$ are defined in Section 4.2. Then*

$$f(\tilde{\nu}(t) - \tilde{b}(t)\omega|w_2; \theta) = \lim_{k\to\infty} f(X(t)|y; \theta; k)$$

and (5.3) holds.

PROOF. (i) In the notation of Sections 3 and 4.1, let

$$\tilde{A}_t = \begin{bmatrix} A \\ \tilde{A}(t) \end{bmatrix}, \qquad \tilde{J}_t = \begin{bmatrix} J & 0 \\ -\tilde{a}(t) & I \end{bmatrix}, \qquad Y_t = \begin{bmatrix} y \\ Y(t) \end{bmatrix}. \tag{5.5}$$

Then

$$\begin{bmatrix} w_1 \\ w_2 \\ \tilde{w}(t) \end{bmatrix} = \tilde{J}_t Y_t = \begin{bmatrix} J_1 A \\ 0 \\ 0 \end{bmatrix} \eta + \begin{bmatrix} J\omega \\ \tilde{\omega}(t) - \tilde{a}(t)\omega \end{bmatrix}.$$

Let $f(w_2, \tilde{w}(t); \theta)$ be the joint density of w_2 and $\tilde{w}(t)$, and let $f(Y_t; \theta; k)$ be the density of Y_t. Then similarly to the proof of Theorem 5.1,

$$f(w_2, \tilde{w}(t); \theta) = c \lim_{k \to \infty} k^{D'/2} f(Y_t; \theta; k),$$

where c is a constant independent of θ, and we can show that c is the same as in (5.1). (5.4) now follows from (5.1), and (5.2) is immediate.

Part (ii) is proved similarly. □

COROLLARY 5.2. (i) *Let* $\tilde{A}_t$ *be defined by* (5.5) *and let* $D'' = \operatorname{rank} \tilde{A}_t$. *If* $D'' > D'$ *then*

$$\lim_{k \to \infty} k^{(D'' - D')/2} f(Y(t) | y; \theta; k) \tag{5.6}$$

exists and is nonzero.

(ii) *Let* $\tilde{B}_t = (A', \tilde{B}(t)')'$ *and let* $D''' = \operatorname{rank} \tilde{B}_t$. *If* $D''' > D'$ *then*

$$\lim_{k \to \infty} k^{(D''' - D')/2} f(X(t) | y; \theta; k)$$

exists and is nonzero.

PROOF. (i) As in the proof of Theorem 5.1(i) we can show that

$$0 < \lim_{k \to \infty} k^{D''/2} f(Y_t; \theta; k) < \infty$$

and (5.6) now follows from (5.1). Part (ii) is proved similarly. □

If the rows of $\tilde{A}(t)$ do not belong to the row space of A, then we cannot estimate $Y(t)$ in the sense of Section 4. By Corollary 5.2(i) this is equivalent to $Y(t) - Y(t|N; k)$ becoming at least partially diffuse as $k \to \infty$. By Corollary 5.2(ii) similar remarks apply to $X(t)$.

For fixed $k > 0$ we can compute $f(y; \theta; k)$ using the Kalman filter as in Schweppe (1965) and Jones (1980). What we need, however, is the limit on the right side of (5.1) and for this we cannot use the ordinary Kalman filter because the operations of filtering and taking the limit cannot be interchanged. In the next section we therefore develop a modification of the Kalman filter to exhibit explicitly the dependence on k. This will provide us with an efficient algorithm to compute the likelihood (5.1).

Similarly, for fixed $k > 0$, we can compute the estimates $Y(t|N; k)$ and $X(t|N; k)$ and their mean-squared errors $S_Y(t|N; k)$ and $S_X(t|N; k)$ using either the fixed point or fixed interval smoothing algorithms described in Chapter 7 of Anderson and Moore (1979). But we need the limit as $k \to \infty$ of the above quantities, and for this we cannot use the ordinary smoothing algorithms because again taking the limit (as $k \to \infty$) and smoothing are not interchangeable operations. In Section 8 we develop modified smoothing algorithms that make explicit the dependence on k. This will give us efficient algorithms to compute (5.2) and (5.3).

6. The modified Kalman filter. For the rest of the paper we assume that Assumption 5.1 holds. For a given integer j, let l be the largest integer for which $t_l \le j$, and define

$$x(t|j; k) = E(x(t)|y(t_1), \dots, y(t_l))$$

and

$$S_x(t|j; k) = \operatorname{Var}(x(t) - x(t|j; k)).$$

We define

$$x(t|0; k) = E(x(t)) = 0, \qquad S_x(t|0; k) = \operatorname{Var}(x(t)),$$

so that

$$x(0|0; k) = 0, \qquad S_x(0|0; k) = k\Phi\Phi' + \Sigma. \tag{6.1}$$

Therefore $x(t|j; k)$ is the best estimate of $x(t)$ given observations until time t_j, and $S_x(t|j; k)$ is its mean-squared error. We define $y(t|j; k)$ and $S_y(t|j; k)$ similarly.

The proofs of the results in Sections 6–8 are often long and require a large number of preliminary results. We have therefore placed all the proofs at the end of the paper in Section 10.

The following lemma makes explicit the dependence of $x(t|j)$, $S_x(t|j)$, and $S_y(t|j)$ on k.

LEMMA 6.1. *For $k > 0$ and $t, j \ge 0$*

$$x(t|j; k) = x^{(0)}(t|j) + O(1/k), \tag{6.2}$$

$$S_x(t|j; k) = kS^{(1)}(t|j) + S^{(0)}(t|j) + O(1/k), \tag{6.3}$$

$$y(t|j; k) = y^{(0)}(t|j) + O(1/k), \tag{6.4}$$

$$S_y(t|j; k) = kS_y^{(1)}(t|j) + S_y^{(0)}(t|j) + O(1/k), \tag{6.5}$$

where $x^{(0)}(t|j)$, $S_x^{(1)}(t|j)$, $S_x^{(0)}(t|j)$, $y^{(0)}(t|j)$, $S_y^{(1)}(t|j)$, and $S_y^{(0)}(t|j)$ do not depend on k.

Let $S_{[y,x]}(t+1|t; k)$ be the conditional variance of $(y(t+1)', x(t+1)')'$ given all observations until time t.

LEMMA 6.2.
(i)

$$S_{[y,x]}(t+1|t;k) = \begin{bmatrix} S_y(t+1|t;k) & H(t+1)S_x(t+1|t;k) \\ S_x(t+1|t;k)H(t+1)' & S_x(t+1|t;k) \end{bmatrix}, \tag{6.6}$$

so that

$$S_{[y,x]}(t+1|t;k) = kS^{(1)}_{[y,x]}(t+1|t) + S^{(0)}_{[y,x]}(t+1|t) + O(1/k). \tag{6.7}$$

(ii) *We can factorize* $S_{[y,x]}(t+1|t;k)$ *as* $L(k)\Lambda(k)L(k)'$, *with* $L(k)$ *a lower triangular matrix and* $\Lambda(k)$ *a diagonal matrix with the expansions*

$$L(k) = L^{(0)} + (1/k)L^{(-1,1)} + (1/k)L^{(-1,0)}(k),$$
$$\Lambda(k) = k\Lambda^{(1)} + \Lambda^{(0)} + O(1/k),$$

where the elements of $L^{(-1,0)}(k)$ *are bounded in absolute value, and the matrices* $L^{(0)}$, $L^{(-1,1)}$, $\Lambda^{(1)}$, $\Lambda^{(0)}$ *are independent of* k. *Furthermore,*

$$L^{(-1,0)}(k)\Lambda^{(1)} = O(1/k).$$

$L^{(0)}$, $L^{(-1,1)}$, $\Lambda^{(1)}$, *and* $\Lambda^{(0)}$ *depend only on* $S_y^{(1)}(t+1|t)$, $S_y^{(0)}(t+1|t)$, $S_x^{(1)}(t+1|t)$, $S_x^{(0)}(t+1|t)$, *and* $H(t+1)$ *and can be computed using Corollary* 10.3.

(iii) *Partition* $L(k)$ *and* $\Lambda(k)$ *as*

$$L(k) = \begin{bmatrix} L_{yy}(k) & 0 \\ L_{yx}(k) & L_{xx}(k) \end{bmatrix}, \qquad \Lambda(k) = \begin{bmatrix} \Lambda_{yy}(k) & 0 \\ 0 & \Lambda_{xx}(k) \end{bmatrix}$$

with $L_{yy}(k)$ *and* $\Lambda_{yy}(k)$ $p \times p$ *matrices and the rest of the submatrices in* $L(k)$ *and* $\Lambda(k)$ *dimensioned conformally. Then*

$$S_x(t+1|t+1;k) = L_{xx}(k)\Lambda_{xx}(k)L_{xx}(k)',$$

so that

$$S_x^{(1)}(t+1|t+1) = L_{xx}^{(0)}\Lambda_{xx}^{(1)}L_{xx}^{(0)\prime} \tag{6.8}$$

and

$$S_x^{(0)}(t+1|t+1) = L_{xx}^{(0)}\Lambda_{xx}^{(0)}L_{xx}^{(0)\prime} + L_{xx}^{(0)}\Lambda_{xx}^{(1)}L_{xx}^{(-1,1)\prime} + L_{xx}^{(-1,1)}\Lambda_{xx}^{(1)}L_{xx}^{(0)\prime}. \tag{6.9}$$

From (ii) $L_{xx}^{(0)}$, $L_{xx}^{(-1,1)}$, $L_{yx}^{(0)}$, $L_{yy}^{(0)}$, $\Lambda_{xx}^{(1)}$, *and* $\Lambda_{xx}^{(0)}$ *depend only on* $S_y^{(1)}(t+1|t)$, $S_y^{(0)}(t+1|t)$, $S_x^{(1)}(t+1|t)$, $S_x^{(0)}(t+1|t)$, *and* $H(t+1)$.

(iv)
$$E\{(x(t+1) - x(t+1|t))|(y(t+1) - y(t+1|t))'\}$$
$$= L_{yx}^{(0)\prime}\tilde{\Lambda}_{yy}L_{yy}^{(0)-1}(y(t+1) - H(t+1)x^{(0)}(t+1|t)) + O(1/k),$$

where $\tilde{\Lambda}_{yy}$ *is a diagonal matrix with* ith *diagonal element* 1 *if the* ith *diagonal element of* Λ_{yy} *is positive and zero otherwise.*

(v) $\operatorname{rank} S_x^{(1)}(t+1|t+1) = \operatorname{rank} S_x^{(1)}(t+1|t) - \operatorname{rank} S_y^{(1)}(t+1|t)$.

Define the innovations $\varepsilon(j;k)$ $(j = 1,\dots,n)$ as $\varepsilon(1;k) = y(1)$ and for $j = 2,\dots,n$

$$\varepsilon(j;k) = y(t_j) - E(y(t_j)|y(t_1),\dots,y(t_{j-1})).$$

Then

$$\varepsilon(j; k) = \varepsilon^{(0)}(j) + O(1/k)$$

with

$$\varepsilon^{(0)}(j) = y(t_j) - H(t_j)x^{(0)}(t_j|t_j - 1)$$

and

$$\mathrm{Var}(\varepsilon(j; k)) = S_y(t_j|t_j - 1; k).$$

We now describe a modification of the ordinary Kalman filter [Anderson and Moore (1979), Chapter 3] in which the dependence on k is made explicit in the state estimates and conditional state covariance matrices.

THEOREM 6.1. *Steps 0–4 below show how to obtain* $x^{(0)}(t|t)$, $S_x^{(0)}(t|t)$, $x^{(0)}(t+1|t)$, $S_x^{(0)}(t+1|t)$, *and* $S_x^{(1)}(t+1|t)$ *for* $t = 0,\dots, N$, *and* $\varepsilon^{(0)}(j)$, $S_y^{(1)}(t_j|t_j - 1)$, *and* $S_y^{(0)}(t_j|t_j - 1)$ *for* $j = 1,\dots, n$.

Step 0 (Initialization)

$$x^{(0)}(0|0) = 0, \qquad S_x^{(1)}(0|0) = \Phi\Phi', \qquad S_x^{(0)}(0|0) = \Sigma. \tag{6.10}$$

Steps 1–4 are repeated for $t = 0,\dots, N - 1$.

Step 1

$$x^{(0)}(t+1|t) = F(t)x^{(0)}(t|t), \tag{6.11}$$

$$\begin{aligned} S_x^{(1)}(t+1|t) &= F(t)S_x^{(1)}(t|t)F(t)', \\ S_x^{(0)}(t+1|t) &= F(t)S_x^{(0)}(t|t)F(t)' + R(t). \end{aligned} \tag{6.12}$$

If $y(t+1)$ *is missing.*

Step 2

$$x^{(0)}(t+1|t+1) = x^{(0)}(t+1|t),$$

$$S_x^{(1)}(t+1|t+1) = S_x^{(1)}(t+1|t), \qquad S_x^{(0)}(t+1|t+1) = S_x^{(0)}(t+1|t).$$

If $y(t+1)$ *is observed* ($t + 1 = t_j$ for some j).

Step 3

$$\varepsilon^{(0)}(j) = y(t_j) - H(t_j)x^{(0)}(t+1|t). \tag{6.13}$$

Now compute $L_{yx}^{(0)}$, $L_{yy}^{(0)}$, $L_{xx}^{(0)}$, $\Lambda_{xx}^{(1)}$, $L_{xx}^{(-1,1)}$, and $\tilde{\Lambda}_{yy}$ as described in Lemma 6.2.

Step 4

$$x^{(0)}(t+1|t+1) = x^{(0)}(t+1|t) + L_{yx}^{(0)\prime}\tilde{\Lambda}_{yy}L_{yy}^{(0)-1}\varepsilon^{(0)}(j). \tag{6.14}$$

$S_x^{(1)}(t+1|t+1)$ and $S_x^{(0)}(t+1|t+1)$ are given by (6.8) and (6.9).

Furthermore,

$$\mathrm{rank}\, S_x^{(1)}(t+1|t+1) = \mathrm{rank}\, S_x^{(1)}(t+1|t) - \mathrm{rank}\, S_y^{(1)}(t+1|t). \tag{6.15}$$

If $S_x^{(1)}(t|t) = 0$ for some t, then $S_x^{(1)}(j|j) = S_x^{(1)}(j+1|j) = 0$ for all $j \geq t$ from (6.12) and (6.15).

From Theorem 6.1 once $S_x^{(1)}(t|t)$ becomes zero it stays zero and the modified Kalman filter reduces to the usual Kalman filter. In applications this will often happen for small t.

The ordinary Kalman filter does not allow a diffuse initial distribution on $x(0)$, and two ways to overcome this have been suggested in the literature. First, one can impose the initial distribution $\eta \sim N(0, kI)$ with k a large positive number and use the ordinary Kalman filter. See, for example, Schweppe [(1973), page 150] and Harvey and Phillips (1979). This method suffers in practice from numerical instabilities as we must often subtract a large number from a large number in order to get a required small number. It is also inexact.

Second, a variant of the Kalman filter known as the information filter [see, for example, Anderson and Moore (1979), Section 6.3] can be used to handle $x(0)$ diffuse when the transition matrices $F(t)$ are nonsingular for $t \geq 0$. The information filter obtains $S_x^{-1}(t|t)$ and $S_x^{-1}(t|t)x(t|t)$, $t = 0, 1, \ldots$ and diffuse initial conditions are imposed by setting $S_x^{-1}(0|0) = 0$. There are several drawbacks to the information filter which makes our approach more appealing. First, the transition matrix $F(t)$ is singular in many cases encountered in practice so its inverse does not exist. An important example of a singular $F(t)$ occurs in a state space representation of an ARIMA (p, d, q) model with $p + d < q + 1$; see, for example, Kohn and Ansley (1984a). Moreover, the conditional covariance matrices $S_x(t|t)$ may be singular as in the ARIMA model of Example 2.1 and thus may not possess the inverses required by the information filter. Finally, even if the information filter can be used, it is computationally more efficient in many applications to work with the $S_x(t|t)$ rather than their inverses. This is particularly so if the dimension of the state vector is large [see Mendel (1971)].

7. Computing the likelihood. When $x(0)$ has a proper distribution, we can use the innovations sequence obtained from the ordinary Kalman filter to efficiently compute the likelihood of a state space model [see, for example, Schweppe (1965) and Jones (1980)]. We similarly use the modified Kalman filter of Section 6 to compute the likelihood (5.1). Note that all quantities in Theorem 7.1 below are readily obtained from the recursions of Theorem 6.1.

THEOREM 7.1. *Define $K_y = \{t_j, j = 1, \ldots, n\}$ so that K_y consists of all those time periods t for which $y(t)$ is observed.*

(i) *Under the assumptions of Sections* 2 *and* 5, *for any $t = t_j$ we can factorize $S_y(t|t-1)$ as $L_j(\theta; k)\Lambda_j(\theta; k)L_j(\theta; k)'$ with*

$$L_j(\theta; k) = L_j^{(0)}(\theta) + O(1/k)$$

and

$$\Lambda_j(\theta; k) = k\Lambda_j^{(1)} + \Lambda_j^{(0)}(\theta) + O(1/k).$$

$L_j^{(0)}(\theta)$ is a lower triangular matrix with 1s *on the diagonal, $L_j^{(0)}(\theta)$ and $\Lambda_j^{(0)}(\theta)$ do not depend on k, and $\Lambda_j^{(1)}$ does not depend on k or θ. We compute $L_j^{(0)}(\theta)$, $\Lambda_j^{(0)}(\theta)$, and $\Lambda_j^{(1)}$ as in Lemma* 6.2(iii).

(ii) *For* $t \in K_y$, *let* $t = t_j$ *and define*

$$\xi(j, \theta) = \left(L_j^{(0)}(\theta)\right)^{-1} \varepsilon^{(0)}(j; \theta)$$

and

$$K_{j0} = \left\{i\colon 1 \le i \le p,\ \lambda_{ij}^{(1)} = 0\right\},$$

where $\lambda_{ij}^{(1)}$ *and* $\lambda_{ij}^{(0)}$ *are the* ith *diagonal elements of* $\Lambda_j^{(1)}(\theta)$ *and* $\Lambda_j^{(0)}(\theta)$, *respectively, and* $\varepsilon^{(0)}(j, \theta)$ *is defined by* (6.13) *in Theorem* 6.1. *Now writing the* ith *element of* $\xi(j)$ *as* ξ_{ij}, *the likelihood* (5.1) *is equal to*

$$c\left\{\prod_{j=1}^{n} \prod_{i \in K_{j0}} \lambda_{ij}^{(0)}\right\}^{-1/2} \exp\left\{-\frac{1}{2}\left(\sum_{j=1}^{n} \sum_{i \in K_{j0}} \xi_{ij}^2/\lambda_{ij}^{(0)}\right)\right\}. \tag{7.1}$$

(iii) *If* $S_y^{(0)}(t|t-1) = H(t)S_x^{(0)}(t|t-1)H(t)' + Q(t)$ *is a positive definite matrix for all* t, *then* $\lambda_{ij}^{(0)} > 0$ *for all* j *and all* $i = 1, \dots, p$. *This happens in particular if* $Q(t)$ *is positive definite for all* t.

8. Modified prediction and smoothing algorithms. In this section we obtain $x^{(0)}(t|N)$, $S_x^{(1)}(t|N)$, $S_x^{(0)}(t|N)$, $y^{(0)}(t|N)$, $S_y^{(1)}(t|N)$, and $S_y^{(0)}(t|N)$ for $t \ge 1$ by extending the smoothing algorithms for the state vector [Anderson and Moore (1979), Chapter 7] to allow for a partially diffuse initial state vector. We will assume throughout that Assumptions 2.1–2.4 and 5.1 hold.

Before giving details of the algorithms we discuss the meaning of our results. Let $y_j^{(0)}(t|N)$ be the jth element of $y^{(0)}(t|N)$ and let $(S_y^{(1)}(t|N))_{jj}$ be the jth diagonal element of $S_y^{(1)}(t|N)$. Suppose $y(t)$ is not observed. Then $y_j(t)$ can be estimated from the data y in the sense of Section 5 if and only if $A_j(t)$ lies in the row space of A, where $A_j(t)$ is the jth row of $A(t)$, and $A(t)$ and A are defined in Section 2. Although it is tedious to directly check whether $A_j(t)$ lies in the row space of A, Theorem 6.1 together with the algorithms below allow us to check it easily and automatically because by Theorem 5.2 and Corollary 5.2, $y_j(t)$ can be estimated from the data in the sense of Section 5 if and only if $(S_y^{(1)}(t|N))_{jj} = 0$. Similar remarks apply to the estimation of $x(t)$ from the data.

8.1. Prediction. For $t > N$ we can use Steps 1 and 2 of Theorem 6.1 to obtain $x^{(0)}(t|N)$, $S_x^{(1)}(t|N)$, and $S_x^{(0)}(t|N)$, as follows.

THEOREM 8.1. *For* $t = N, N+1 \dots$

$$x^{(0)}(t+1|N) = F(t)x^{(0)}(t|N), \qquad S_x^{(1)}(t+1|N) = F(t)S_x^{(1)}(t|N)F(t)',$$

$$S_x^{(0)}(t+1|N) = F(t)S_x^{(0)}(t|N)F(t)' + R(t).$$

The modified fixed interval, fixed point and fixed lag smoothing algorithms described below apply to $t < N$.

8.2. Modified fixed interval smoothing. Let $S_{[t+1,t]}(k)$ be the conditional variance–covariance matrix of $(x(t+1)', x(t)')'$ given all available observations until time t.

LEMMA 8.1.
(i)

$$S_{[t+1,t]}(k)=\begin{bmatrix} S_x(t+1|t;k) & F(t)S_x(t|t;k) \\ S_x(t|t;k)F(t)' & S_x(t|t;k)\end{bmatrix}. \tag{8.1}$$

(ii) *We can factorize* $S_{[t+1,t]}(k)$ *as* $L(k)\Lambda(k)L(k)'$, *where* $L(k)$ *is a lower triangular matrix and* $\Lambda(k)$ *a diagonal matrix.* $L(k)$ *and* $\Lambda(k)$ *can be expanded as*

$$L(k)=L^{(0)}+(1/k)L^{(-1,1)}+(1/k)L^{(-1,0)}(k),$$
$$\Lambda(k)=k\Lambda^{(1)}+\Lambda^{(0)}+O(1/k)$$

with the elements of $L^{(-1,0)}(k)$ *bounded in absolute value, and the matrices* $L^{(0)}$, $L^{(-1,1)}$, $\Lambda^{(1)}$, *and* $\Lambda^{(0)}$ *independent of* k. *Furthermore,*

$$L^{(-1,0)}(k)\Lambda^{(1)}=O(1/k).$$

$L^{(0)}$, $L^{(-1,1)}$, $\Lambda^{(1)}$, *and* $\Lambda^{(0)}$ *depend only on* $S_x^{(1)}(t+1|t)$, $S_x^{(1)}(t|t)$, $S_x^{(0)}(t|t)$, *and* $F(t)$, *and can be computed from* (8.1) *using Corollary* 10.3.

(iii) *Partition* $L(k)$ *and* $\Lambda(k)$ *as*

$$L(k)=\begin{bmatrix} L_{11}(k) & 0 \\ L_{21}(k) & L_{21}(k)\end{bmatrix},\qquad \Lambda(k)=\begin{bmatrix}\Lambda_{11}(k) & 0 \\ 0 & \Lambda_{22}(k)\end{bmatrix}$$

with $L_{11}(k)$ *and* $\Lambda_{11}(k)$ $q\times q$ *matrices and the rest of the submatrices partitioned conformally. Then,*

$$\begin{aligned}&E\{(x(t)-x(t|t;k))|(x(t+1)-x(t+1|t;k))\}\\ &\quad=C(t;k)(x(t+1)-x(t+1|t;k)),\end{aligned} \tag{8.2}$$

where

$$C(t;k)=L_{21}(k)\tilde{\Lambda}_{11}L_{11}^{-1}(k). \tag{8.3}$$

$\tilde{\Lambda}_{11}$ *is a diagonal matrix with* ith *element* 1 *if the* ith *element of* $\Lambda_{11}(k)$ *is positive and zero otherwise.*

We can write

$$C(t;k)=C^{(0)}(t)+(1/k)C^{(-1,1)}(t)+(1/k)C^{(-1,0)}(t;k)+O(1/k^2),$$

where $C^{(0)}(t)$ *and* $C^{(-1,1)}(t)$ *are independent of* k *and* $C^{(-1,0)}(t;k)$ *has elements that are bounded in absolute value. Furthermore,*

$$C^{(0)}(t)=L_{21}^{(0)}\tilde{\Lambda}_{11}\left(L_{11}^{(0)}\right)^{-1}, \tag{8.4}$$

$$C^{(-1,1)}(t)=L_{21}^{(-1,1)}\tilde{\Lambda}_{11}\left(L_{11}^{(0)}\right)^{-1}-L_{21}^{(0)}\tilde{\Lambda}_{11}\left(L_{11}^{(0)}\right)^{-1}L_{11}^{(-1,1)}\left(L_{11}^{(0)}\right)^{-1}, \tag{8.5}$$

$$\begin{aligned}C^{(-1,0)}(t;k)&=L_{21}^{(-1,0)}(k)\tilde{\Lambda}_{11}\left(L_{11}^{(0)}\right)^{-1}\\ &\quad-L_{21}^{(0)}\tilde{\Lambda}_{11}\left(L_{11}^{(0)}\right)^{-1}L_{11}^{(-1,0)}(k)\left(L_{11}^{(0)}\right)^{-1},\end{aligned} \tag{8.6}$$

and

$$C^{(-1,0)}(t;k)S_x^{(1)}(t+1|t)=O(1/k). \tag{8.7}$$

THEOREM 8.2 (Modified fixed interval smoothing). *We consider the state space model* (2.1) *with Assumptions* 2.1–2.4 *and* 5.1 *holding.*

For $t = N-1,\dots,1$

$$x^{(0)}(t|N) = x^{(0)}(t|t) + C^{(0)}(t)\big(x^{(0)}(t+1|N) - x^{(0)}(t+1|t)\big), \tag{8.8}$$

$$S_x^{(1)}(t|N) = S_x^{(1)}(t|t) - C^{(0)}(t)\big(S_x^{(1)}(t+1|t) - S_x^{(1)}(t+1|N)\big)C^{(0)}(t)', \tag{8.9}$$

$$\begin{aligned} S_x^{(0)}(t|N) = S_x^{(0)}(t|t) &- C^{(0)}(t)\big(S_x^{(0)}(t+1|t) - S_x^{(0)}(t+1|N)\big)C^{(0)}(t)' \\ &- C^{(0)}(t)\big(S_x^{(1)}(t+1|t) - S_x^{(1)}(t+1|N)\big)C^{(-1,1)}(t)' \\ &- C^{(-1,1)}(t)\big(S_x^{(1)}(t+1|t) - S_x^{(1)}(t+1|N)\big)C^{(0)}(t)'. \end{aligned} \tag{8.10}$$

$S_x^{(1)}(t+1|t)$, $S_x^{(0)}(t+1|t)$, $S_x^{(1)}(t|t)$, *and* $S_x^{(0)}(t|t)$ *are computed by Theorem* 6.1, $C^{(0)}(t)$ *is obtained from* (8.4) *and* $C^{(-1,1)}(t)$ *is obtained from* (8.5).

8.3. Fixed point smoothing. The fixed interval smoothing algorithm systematically obtains $x^{(0)}(t|N)$, $S_x^{(1)}(t|N)$, and $S_x^{(0)}(t|N)$ for $t = N-1,\dots,1$. If we need these quantities for just a small number of values of t, then using fixed point smoothing may be faster. For a detailed implementation of this algorithm in the scalar observation case ($p = 1$) see Kohn and Ansley (1984a).

The fixed point smoothing algorithm [Anderson and Moore (1979), Section 7.2] obtains $x(t|N)$ and $S_x(t|N)$ by recursively updating $x(t|t+j)$ and $S_x(t|t+j)$ for $j = 0,\dots, N-t$, with starting values $x(t|t)$ and $S_x(t|t)$ at $j = 0$ being obtained from the Kalman filter.

Following Anderson and Moore [(1979), Section 7.2], we use the augmented state vector $z(t+j) = (x(t+j)', x(t)')'$ for $j \ge 0$ to obtain an augmented set of state space equations

$$y(t+j) = (H(t+j), 0)z(t+j) + v(t+j), \tag{8.11}$$

$$z(t+j+1) = \begin{bmatrix} F(t+j) & 0 \\ 0 & I \end{bmatrix} z(t+j) + \begin{bmatrix} \begin{bmatrix} u(t+j) \\ 0 \end{bmatrix} \end{bmatrix} \tag{8.12}$$

We can apply Theorem 6.1 to (8.11) and (8.12) to obtain $z^{(0)}(t+j|t+j)$, $S_z^{(1)}(t+j|t+j)$, and $S_z^{(0)}(t+j|t+j)$ and so in particular $x^{(0)}(t|t+j)$, $S_x^{(1)}(t|t+j)$, and $S_x^{(0)}(t|t+j)$ for $j = 1,\dots, N-t$. Starting values for $j = 0$ are given by $z^{(0)}(t|t)$, $S_z^{(1)}(t|t)$, and $S_z^{(0)}(t|t)$, which are obtained immediately from $x^{(0)}(t|t)$, $S_x^{(1)}(t|t)$, and $S_x^{(0)}(t|t)$.

8.4. Fixed lag smoothing. Fixed lag smoothing is a recursive procedure for obtaining $x(t|t+l)$ and $S_x(t|t+l)$ for fixed $l > 0$. To derive the algorithm we augment the state vector as in Anderson and Moore [(1979), Section 7.3], and then apply the modified Kalman filter (Theorem 6.1).

8.5. Prediction and interpolation of the dependent variable. The following lemma shows that in two special but important cases the algorithms described in Sections 8.1–8.4 above enable us to obtain immediately the expectation of any missing $y(t)$ conditional on y, and also the mean-squared error of the estimate.

LEMMA 8.2. (i) *If* $y(t)$ *is a subvector of* $x(t)$ *for all* t, *then* $y^{(0)}(t|N)$, $S_y^{(1)}(t|N)$, *and* $S_y^{(0)}(t|N)$ *are obtained automatically from Sections* 8.1–8.4 *above.*

(ii) *If* $v(t)$ *is identically zero, then*

$$y(t|N) = H(t)x(t|N) \quad and \quad S_y(t|N) = H(t)S_x(t|N)H(t)'.$$

The proof is immediate. □

One particularly important application of Lemma 8.2 is to the ARIMA model because here $y(t)$ is a subvector of $x(t)$; in Example 2.1, for instance, $y(t)$ is the first element of $x(t)$.

We now present a more general approach for obtaining $y(t|N)$ and $s_y(t|N)$ for missing $y(t)$. For each t, define the augmented state vector $z(t) = (y(t)', x(t)')'$, so that $y(t) = (I, 0)z(t)$, and

$$z(t+1) = \begin{bmatrix} 0 & H(t+1)F(t) \\ 0 & F(t) \end{bmatrix} z(t) + \begin{bmatrix} H(t+1)u(t) + v(t) \\ u(t) \end{bmatrix}$$

become our new state equations. We can now apply the modified Kalman filter and the algorithms in Sections 8.1–8.4 to obtain $z(t|N)$ and $S_z(t|N)$ and so in particular $y(t|N)$ and $S_y(t|N)$ for the missing $y(t)$. Because in many applications the dimension of $x(t)$ will be much greater than that of $y(t)$, using the augmented state vector will not increase computing time significantly.

8.6. Partially missing or aggregated observations. So far we have assumed (Assumption 2.4) that for each t we either observe all of $y(t)$ or none of it. More generally, all our results continue to hold if for each t we observe not $y(t)$ but $z(t) = T(t)y(t)$ where $T(t)$ is known for each t. Thus $z(t)$ may consist of just some of the elements of $y(t)$ or an aggregate of the elements of $y(t)$. See Kohn and Ansley (1983a, b) for details.

9. Continuous time state space models observed at discrete time points. We now briefly describe how our results continue to hold when the state vector $x(t)$ is generated by the continuous time state transition equation

$$dx(t)/dt = F(t;\theta)x(t) + g(t;\theta)\,dW(t)/dt, \qquad t \geq 0, \tag{9.1}$$

and the observation equation

$$y(t_i) = H(t_i;\theta)x(t_i) + v(i), \qquad i = 1, \ldots, n.$$

We assume that,

ASSUMPTION 9.1. (i) The observation times are ordered as $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$.

(ii) The observations $y(t_i)$ are $p \times 1$ and the state vector $x(t)$ is $q \times 1$.

(iii) $W(t)$ is a zero mean Wiener process with $W(0) = 0$ and $\mathrm{Var}(W(1)) = 1$.

(iv) $v(j)$ $(j = 1, \ldots, n)$ is a $p \times 1$ sequence of independent $N(0, Q(j;\theta))$ random variables which is independent of $W(t)$.

(v) $F(t; \theta)$ and $g(t; \theta)$ are $q \times q$ and $q \times 1$, respectively, and are continuous functions of t for each $\theta \in \Theta$. $H(t; \theta)$ is $p \times q$.

(vi) The initial conditions are given by Assumption 2.2.

To use the results in Sections 2–8, we rewrite (9.1) in integral form. Let $\chi(t; \theta)$ be the $q \times q$ matrix solution of $d\chi(t)/dt = F(t)\chi(t)$ with $\chi(0) = I_q$. By Hochstadt [(1964), page 79], $\chi(t)$ exists and is unique. Following Hochstadt [(1964), Section 2.7], let $\phi(t, s; \theta) = \chi(t; \theta)\chi(s; \theta)^{-1}$. For convenience we often omit to indicate dependence on θ below. Then for $s < t$ we can rewrite (9.1) as

$$x(t) = \phi(t, s)x(s) + u(t, s), \tag{9.2}$$

where

$$u(t, s) = \int_s^t \phi(t, t')b(t')\, dW(t').$$

Let $R(t, s; \theta) = \operatorname{Var}(u(t, s))$. Taking $s = t_i$ and $t = t_{i+1}$ in (9.2) we obtain

$$x(t_{i+1}) = \phi(t_{i+1}, t_i)x(t_i) + u(t_{i+1}, t_i),$$

which corresponds to the state transition equation in (2.1).

Now let $A(t) = H(t)\phi(t,0)\Phi$, $\nu(t) = \phi(t,0)\zeta + u(t,0)$, and $\omega(t_i) = H(t_i)\nu(t_i) + v(i)$ $(i = 1,\dots, n)$. Then $y(t_i) = A(t_i)\eta + \omega(t_i)$ and

$$y = A\eta + \omega$$

with y, ω, and A being defined as in Section 2. If we assume that $A(t)$ is independent of θ for all t, then we can define the likelihood of y as in Section 3.

Now let $B(t) = \phi(t,0)\Phi$ so that

$$x(t) = B(t)\eta + \nu(t),$$

and we can define estimates of the elements of $x(t)$ as in Section 4.

Efficient algorithms for computing the likelihood and the estimates of the state vectors at the observation times $t_1, \dots, t_n$ are again obtained by taking $\eta \sim N(0, kI)$ and obtaining analogues of the filtering and smoothing algorithms of Sections 6–8. To obtain estimates of $x(t)$ for $t_i < t < t_{i+1}$, some further modification of the fixed interval smoothing algorithm is required, as described in Weinert et al. (1980) and Kohn and Ansley (1984b). We omit details.

We conclude by giving examples of two applications of the results sketched out above.

EXAMPLE 9.1. We consider estimation of a continuous time second-order autoregressive process observed at discrete points in time. An application of such models is given in Jones (1983).

Suppose we have the scalar observations $y(t_i) = f(t_i)$ with $f(t)$ generated by the second-order differential equations

$$d^2f(t)/dt^2 = b_1\, df(t)/dt + b_0 f(t) + \sigma\, dW(t)/dt, \qquad t \geq 0, \tag{9.3}$$

so that the unknown parameters are $\theta = (b_0, b_1, \sigma^2)$. Then we can rewrite the observation equation and (9.3) in state space form as

$$y(t_i) = Hx(t_i), \qquad dx(t)/dt = F(\theta)x(t) + g(\theta)\, dW(t)/dt$$

with $x(t) = (f(t), df(t)/dt)'$, $H = (1,0)$, $g(\theta) = (0, \sigma)'$, and

$$F(\theta) = \begin{bmatrix} 0 & 1 \\ b_0 & b_1 \end{bmatrix}.$$

If we take the distribution of $x(0)$ as unknown, then we can form the likelihood of our observations and obtain estimates of unobserved $f(t)$ as outlined above.

EXAMPLE 9.2. Cubic spline smoothing. Consider the scalar model

$$y(t_i) = f(t_i) + v(i), \tag{9.4}$$

$$d^2 f(t)/dt^2 = \sigma\sqrt{\mu}\, dW(t)/dt, \tag{9.5}$$

where $v(i)$ is an $N(0, \sigma^2)$ independent sequence and is independent of $W(t)$; $\mu > 0$. As in Example 9.1, we can write (9.4) and (9.5) in state space form. If we take $x(0) \sim N(0, kI)$ then for given μ and σ^2 we have by Wahba (1978) that

$$\lim_{k \to \infty} E(x_1(t) | y) \tag{9.6}$$

is the optimal cubic spline smoothing the data $y(t_1), \ldots, y(t_n)$ with smoothness parameter $1/\mu$. (9.6) can be computed efficiently using our modified filtering and smoothing algorithms.

A general treatment of smoothing and interpolating splines using our modified algorithms is given by Kohn and Ansley (1984b).

10. Proofs

10.1. Preliminary Results. We set out a number of results on the Cholesky factorization of a positive semidefinite matrix indexed by a scale parameter k, and characterize conditional variances and expectations of Gaussian random vectors whose unconditional covariance matrices take this form. These results are needed to prove the results in Sections 6–8.

Much of our work involves matrices satisfying the following condition:

CONDITION I. $\Omega(k)$ is a $q \times q$ matrix of the form

$$\Omega(k) = k\Omega^{(1)} + \Omega^{(0)},$$

where $k > 0$, and $\Omega^{(1)}$ and $\Omega^{(0)}$ are positive semidefinite matrices independent of k.

The following result will be used frequently. For a proof see Kohn and Ansley (1983a).

LEMMA 10.1. *Suppose Ω is a positive semidefinite matrix. Then there exists a lower triangular matrix L with* 1s *on the diagonal and a diagonal matrix Λ with nonnegative elements such that $\Omega = L\Lambda L'$. Further, if the* ith *diagonal element of Λ is zero, then the* ith *column of L is zero except for the unit diagonal element. The number of nonzero elements of Λ is equal to the rank of Ω.*

We now extend Lemma 10.1 to characterize the factorization of matrices satisfying Condition I.

THEOREM 10.1. *Suppose that $\Omega(k)$ is a $q \times q$ matrix satisfying Condition I. Then for $k > 0$, $\Omega(k)$ is a positive semidefinite matrix and can be factorized as*

$$\Omega(k) = L(k)\Lambda(k)L(k)', \tag{10.1}$$

where

(i) *$\Lambda(k)$ is a diagonal matrix and $L(k)$ a lower triangular matrix with* 1s *on the diagonal.*

(ii) *We can write*

$$L(k) = L^{(0)} + (1/k)L^{(-1)}(k),$$
$$\Lambda(k) = k\Lambda^{(1)} + \Lambda^{(0)} + (1/k)\Lambda^{(-1)}(k),$$

where $L^{(0)}$, $\Lambda^{(1)}$, and $\Lambda^{(0)}$ do not depend on k, and the elements of $L^{(-1)}(k)$ and $\Lambda^{(-1)}(k)$ are bounded in absolute value.

(iii) *Let the* ith *diagonal elements of $\Lambda^{(1)}$, $\Lambda^{(0)}$, and $\Lambda^{(-1)}(k)$ be $\lambda_i^{(1)}$, $\lambda_i^{(0)}$, and $\lambda_i^{(-1)}(k)$, respectively. Then $\lambda_i^{(1)} \geq 0$ and $\lambda_i^{(0)} \geq 0$ if $\lambda_i^{(1)} = 0$; $\lambda_i^{(-1)}(k) = 0$ whenever both $\lambda_i^{(1)} = 0$ and $\lambda_i^{(0)} = 0$.*

(iv) *Let the* ijth *element of $L(k)$ be $l_{ij}(k)$. Then if $\lambda_j^{(1)} > 0$ we can write*

$$l_{ij}(k) = (1/k)l_{ij}^{(-1)} + O(1/k^2),$$

where $l_{ij}^{(-1)}$ does not depend on k.

(v) *The elements of $\Lambda^{(1)}$ depend on $\Omega^{(1)}$ but not $\Omega^{(0)}$.*

PROOF. It is immediate that $\Omega(k)$ is positive semidefinite for all $k \geq 0$. The factorization (10.1) now follows directly from Lemma 10.1.

Let $\omega_{ij}^{(1)}$, $\omega_{ij}^{(0)}$ be the ijth elements of $\Omega^{(1)}$ and $\Omega^{(0)}$, respectively. We first consider the case $\Omega^{(1)}$ diagonal and show that:

(i) For $i = 1, \dots, q$

$$\lambda_i(k) = k\lambda_i^{(1)} + \lambda_i^{(0)} + (1/k)\lambda_i^{(-1)}(k), \tag{10.2}$$

where $\lambda_i^{(1)}$ and $\lambda_i^{(0)}$ do not depend on k, $\lambda_i^{(1)} \geq 0$, and $\lambda_i^{(0)} \geq 0$ if $\lambda_i^{(1)} = 0$, and $|\lambda_i^{(-1)}(k)|$ is bounded. If $\lambda_i^{(1)} = \lambda_i^{(0)} = 0$ then $\lambda_i^{(-1)}(k) = 0$.

(ii) For $i > 1$ and $j = 1, \dots, i - 1$

$$l_{ij}(k) = l_{ij}^{(0)} + O(1/k), \tag{10.3}$$

where $l_{ij}^{(0)}$ is independent of k. If $\lambda_j^{(1)} > 0$, then $l_{ij}^{(0)} = 0$ and

$$l_{ij}(k) = (1/k)l_{ij}^{(-1)} + O(1/k^2), \tag{10.4}$$

where $l_{ij}^{(-1)}$ does not depend on k.

Note first that

$$\lambda_1(k) = k\omega_{11}^{(1)} + \omega_{11}^{(0)},$$

so that $\lambda_1^{(1)} = \omega_{11}^{(1)}$, $\lambda_1^{(0)} = \omega_{11}^{(0)}$, and $\lambda_1^{(-1)}(k) = 0$. Hence (10.3) holds for $i = 1$.

We now proceed by induction. For convenience of notation, we omit the dependence on k.

Assume (10.2)–(10.4) hold for $i < r$, $j < i$, and for $i = r$, $j < s < r$.

Define $H_1(s) = \{h\colon 1 \le h < s, \lambda_h^{(1)} > 0\}$ and $H_0(s) = \{h\colon 1 \le h < s, \lambda_h^{(1)} = 0\}$. If $\lambda_s = 0$ then $l_{rs} = 0$ by Lemma 10.1. Otherwise, we can construct l_{rs} by

$$(10.5)\qquad \begin{aligned} l_{rs} = \Big[\omega_{rs} - \sum_{H_1(s)} &\big((1/k)l_{rh}^{(-1)} + O(1/k)\big)\big((1/k)l_{sh}^{(-1)} + O(1/k)\big) \\ &\times\big(k\lambda_h^{(1)} + \lambda_h^{(0)} + O(1/k)\big) \\ &- \sum_{H_0(s)} \big(l_{rh}^{(0)} + O(1/k)\big)\big(l_{sh}^{(0)} + O(1/k)\big)\big(\lambda_h^{(0)} + O(1/k)\big)\Big] \\ &\Big/\big(k\lambda_s^{(1)} + \lambda_s^{(0)} + O(1/k)\big). \end{aligned}$$

If $\lambda_s^{(1)} > 0$ we can write

$$(10.6)\qquad \big(k\lambda_s^{(1)} + \lambda_s^{(0)} + O(1/k)\big)^{-1} = \big(1/k\lambda_s^{(1)}\big)\big(1 - \lambda_s^{(0)}/\big(k\lambda_s^{(1)}\big) + O(1/k^2)\big)$$

and obtain (10.4) with

$$l_{rs}^{(-1)} = \Big(\omega_{rs}^{(0)} - \sum_{H_0(s)} l_{rh}^{(0)} l_{sh}^{(0)} \lambda_h^{(0)}\Big)\Big/\lambda_s^{(1)}.$$

Similarly, if $\lambda_s^{(1)} = 0$ (10.5) holds with

$$l_{rs}^{(0)} = \Big(\omega_{rs}^{(0)} - \sum_{H_0(s)} l_{rh}^{(0)} l_{sh}^{(0)} \lambda_h^{(0)}\Big)\Big/\lambda_s^{(0)}.$$

Now

$$\begin{aligned} \lambda_r &= k\omega_{rr}^{(1)} + \omega_{rr}^{(0)} - \sum_{H_1(r)} \big((1/k)l_{rh}^{(-1)} + O(1/k)\big)^2\big(k\lambda_h^{(1)} + \lambda_h^{(0)} + O(1/k)\big) \\ &\quad - \sum_{H_0(r)} \big(l_{rh}^{(0)} + O(1/k)\big)^2\big(\lambda_h^{(0)} + O(1/k)\big) \\ &= k\omega_{rr}^{(1)} + \Big(\omega_{rr}^{(0)} - \sum_{H_0(r)} \big\{l_{rh}^{(0)}\big\}^2 \lambda_h^{(0)}\Big) + O(1/k), \end{aligned}$$

so that (10.3) holds with $\lambda_r^{(1)} = \omega_{rr}^{(1)}$ and

$$\lambda_r^{(0)} = \omega_{rr}^{(0)} - \sum_{H_0(r)} \big(l_{rh}^{(0)}\big)^2 \lambda_r^{(0)}.$$

Note that $\lambda_r(k) \ge 0$ for $k > 0$, so that $\lambda_r^{(1)} \ge 0$ and $\lambda_r^{(0)} \ge 0$ if $\lambda_r^{(1)} = 0$.

As shown in Lemma 10.2 below, Condition I implies that $\lambda_i^{(-1)} = 0$ if $\lambda_i^{(1)} = \lambda_i^{(0)} = 0$. Finally, the rank of $\Omega^{(1)}$ is equal to the number of nonzero $\omega_{rr}^{(1)}$ elements, and hence to the number of nonzero $\lambda_r^{(1)}$ elements.

Part (v) follows because $\lambda_r^{(1)} = \omega_{rr}^{(1)}$ for all r.

Now suppose $\Omega^{(1)}$ is not diagonal. From Lemma 10.1 we can write $\Omega^{(1)} = BDB'$, where D is diagonal and B is a lower triangular matrix with 1s on the diagonal. Note that the number of positive diagonal elements of D equals the rank of $\Omega^{(1)}$. The results proved above hold for $B^{-1}\Omega(k)(B^{-1})'$ and hence for $\Omega(k)$ also. □

LEMMA 10.2. *Suppose the conditions of Theorem* 10.1 *hold. Then, for all i*

$$\lambda_i^{(-1)}(k) = 0 \quad \text{if } \lambda_i^{(1)} = \lambda_i^{(0)} = 0.$$

PROOF. By Theorem 10.1, $\Omega(k)$ is a positive semidefinite matrix so that we can regard it as the variance covariance matrix of a q dimensional zero mean vector random variable z. Let $M(k) = L^{-1}(k)$. Then it is not difficult to check that we can write

$$M(k) = M^{(0)} + \frac{1}{k}M^{(-1)}(k)$$

with the elements of $M^{(-1)}(k)$ bounded in absolute value, and $M^{(0)}$ independent of k. Furthermore, $M^{(0)}$ is lower triangular with 1s on the diagonal and $M^{(-1)}(k)$ is lower triangular with diagonal elements zero. If we put $\varepsilon = M(k)z$, then the elements of ε are the innovations of z and $\lambda_i(k) = \mathrm{Var}(\varepsilon_i)$, where ε_i is the ith element of ε.

Let $\alpha(k)'$, β', and $\gamma(k)'$ be the ith rows of $M(k)$, $M^{(0)}$ and $M^{(-1)}(k)$, respectively. Then

$$\lambda_i(k) = \alpha(k)'\Omega(k)\alpha(k) = k\beta'\Omega^{(1)}\beta + 2\beta'\Omega^{(1)}\gamma + \beta'\Omega^{(0)}\beta + O(1/k).$$

If $\lambda_i^{(1)} = 0$ it follows that $\beta'\Omega^{(1)}\beta = 0$ and hence $\beta'\Omega^{(1)}\gamma = 0$. If, in addition, $\lambda_i^{(0)} = 0$ then $\beta'\Omega^{(0)}\beta = 0$, so that $\beta'\Omega(k)\beta = 0$. Now $\beta'z = -\gamma(k)'z/k + \varepsilon_i$, and because elements $i, i+1, \dots, q$ of $\gamma(k)$ are zero and ε_i is an innovation, $\gamma(k)'z$ is independent of ε_i. Thus

$$\lambda_i(k) = \mathrm{Var}(\varepsilon_i) = \mathrm{Var}(\beta'z) - \mathrm{Var}(\gamma(k)'z)/k^2 \le \mathrm{Var}(\beta'z) = 0.$$

The result follows. □

REMARK 10.1. In the applications below z is a vector random variable with a variance–covariance matrix satisfying Condition I. Then for each i, $\lambda_i(k)$ is the ith innovation variance, and Lemma 10.2 tells us that either $\liminf \lambda_i(k) > 0$ or $\lambda_i(k) = 0$ for all $k > 0$. That is, either the ith innovation variance is bounded away from zero for all $k > 0$, or the ith innovation is identically zero.

REMARK 10.2. In Theorem 10.1 it is possible to show that $\lambda_i^{(0)} \ge 0$ for all i, including values of i for which $\lambda_i^{(1)} > 0$. Because we do not need this property for subsequent results, and its proof is quite long, we have chosen to omit it.

COROLLARY 10.1. *Suppose the conditions of Theorem* 10.1 *hold, and that in addition,* $\Omega^{(0)}$ *is positive definite. Then, in the notation of Theorem* 10.1, $\lambda_i^{(0)} > 0$ *for* $i = 1$ *to* q.

PROOF. Suppose first that $\Omega^{(1)}$ is diagonal. Then, from the proof of Theorem 10.1, $\lambda_1^{(0)} = \omega_{11}^{(0)} > 0$.

For $i = 2, \dots, q$, let $\Omega_i^{(0)}$ and $\Omega_i^{(1)}$ be, respectively, the upper left $(i-1) \times (i-1)$ submatrices of $\Omega^{(0)}$ and $\Omega^{(1)}$. Let $\alpha(k)$ be the column vector consisting of the first $(i-1)$ elements of the ith row of $\Omega(k)$. Then $\alpha(k)$ consists of the first

$(i-1)$ elements of the ith row of $\Omega^{(0)}$ because $\Omega^{(1)}$ is diagonal. We define the random vectors z and ε as in Lemma 10.2.

Then for $i \geq 2$,

$$\lambda_i(k) = \mathrm{Var}(\varepsilon_i) = k\omega_{ii}^{(1)} + \omega_{ii}^{(0)} - \alpha'\left(k\Omega_i^{(1)} + \Omega_i^{(0)}\right)^{-1}\alpha,$$

and because $\lambda_i^{(1)} = \omega_{ii}^{(1)}$,

$$\lambda_i^{(0)} = \omega_{ii}^{(0)} - \alpha'\left(k\Omega_i^{(1)} + \Omega_i^{(0)}\right)^{-1}\alpha + O(1/k) \geq \omega_{ii}^{(0)} - \alpha'\left\{\Omega_i^{(0)}\right\}^{-1}\alpha + O(1/k) > 0$$

for large k. The inequality $\omega_{ii}^{(0)} - \alpha'\{\Omega_i^{(0)}\}^{-1}\alpha > 0$ holds because $\Omega^{(0)}$ is positive definite.

To complete the proof we note that we can reduce the case of general $\Omega^{(1)}$ to the diagonal case as in the proof of Theorem 10.1. □

The results of Theorem 10.1 are now restated in a more detailed form in the following corollary, which can be used to form an efficient computational algorithm.

COROLLARY 10.2. *The elements of the factorization established by Theorem 10.1 are given in general by the following formulae*:

$$\lambda_1^{(1)} = \omega_{11}^{(1)}, \qquad \lambda_1^{(0)} = \omega_{11}^{(0)}. \tag{10.7}$$

For $i > 1$

$$\lambda_i^{(1)} = \omega_{ii}^{(1)} - \sum_{h=1}^{i-1} \left(l_{ih}^{(0)}\right)^2 \lambda_h^{(1)}, \tag{10.8}$$

$$\lambda_i^{(0)} = \omega_{ii}^{(0)} - \sum_{h=1}^{i-1} \left(2 l_{ih}^{(0)} l_{ih}^{(-1)} \lambda_h^{(1)} + \left(l_{ih}^{(0)}\right)^2 \lambda_h^{(0)}\right). \tag{10.9}$$

If $\lambda_j^{(1)} > 0$ *then*

$$l_{ij}^{(0)} = \left(\omega_{ij}^{(1)} - \sum_{h=1}^{j-1} l_{ih}^{(0)} l_{jh}^{(0)} \lambda_h^{(1)}\right) \Big/ \lambda_j^{(1)}, \tag{10.10}$$

$$\begin{aligned} l_{ij}^{(-1)} = \frac{1}{\lambda_j^{(1)}} \Bigg\{ \omega_{ij}^{(0)} &- \sum_{H_1(j)} \left[\left(l_{ih}^{(0)} l_{jh}^{(-1)} + l_{ih}^{(-1)} l_{jh}^{(0)}\right)\lambda_h^{(1)}\right] \\ &- \sum_{h=1}^{j-1} l_{ih}^{(0)} l_{jh}^{(0)} \lambda_h^{(0)} - l_{ij}^{(0)} \lambda_j^{(0)} \Bigg\}, \end{aligned} \tag{10.11}$$

and if $\lambda_j^{(1)} = 0$, $\lambda_j^{(0)} > 0$ *then*

$$l_{ij}^{(0)} = \frac{1}{\lambda_j^{(0)}} \left\{ \omega_{ij}^{(0)} - \sum_{h=1}^{j-1} \left[\left(l_{ih}^{(0)} l_{jh}^{(-1)} + l_{ih}^{(-1)} l_{jh}^{(0)}\right)\lambda_h^{(1)} - l_{ih}^{(0)} l_{jh}^{(0)} \lambda_h^{(0)}\right] \right\}. \tag{10.12}$$

Proof. (10.7) follows from the proof of Theorem 10.1. Defining $H_1(j)$ and $H_0(j)$ as in the proof of Theorem 10.1, we can write for $i > j$ and $\lambda_j > 0$

$$l_{ij} = \Big\{ k\omega_{ij}^{(1)} + \omega_{ij}^{(0)} + O(1/k) - \sum_{H_1(j)} \Big[\big(l_{ih}^{(0)} + (1/k) l_{ih}^{(-1)} + O(1/k^2) \big) \big(l_{jh}^{(0)} + (1/k) l_{jh}^{(-1)} + O(1/k^2) \big) \times \big(k\lambda_h^{(1)} + \lambda_h^{(0)} + O(1/k) \big) \Big] - \sum_{H_0(j)} \big(l_{ih}^{(0)} + O(1/k) \big) \big(l_{jh}^{(0)} + O(1/k) \big) \big(\lambda_h^{(0)} + O(1/k) \big) \Big\} \Big/ \big(k\lambda_j^{(1)} + \lambda_j^{(0)} + O(1/k) \big). \tag{10.13}$$

If $\lambda_j^{(1)} > 0$ we can write $\{k\lambda_j^{(1)} + \lambda_j^{(0)} + O(1/k)\}^{-1}$ as in (10.6), and (10.10) and (10.11) follow by substitution in (10.13).

If $\lambda_j^{(1)} = 0$ we can substitute in (10.13) to obtain

$$l_{ij} = k \Big[\omega_{ij}^{(1)} - \sum_{H_1(j)} l_{ih}^{(0)} l_{jh}^{(0)} \lambda_h^{(1)} \Big] \Big/ \lambda_j^{(0)} + \Big[\omega_{ij}^{(0)} - \sum_{H_1(j)} \big(l_{ih}^{(0)} l_{jh}^{(-1)} + l_{ih}^{(-1)} l_{jh}^{(0)} \big) \lambda_h^{(1)} - \sum_{H_0(j)} l_{ih}^{(0)} l_{jh}^{(0)} \lambda_h^{(0)} \Big] \Big/ \lambda_j^{(0)} + O(1/k).$$

The first term on the right-hand side is zero by Part (ii) of Theorem 10.1, and the second term gives (10.12).

(10.8) and (10.9) follow similarly. □

Corollary 10.3. *Suppose $\Omega(k)$ satisfies Condition I, and has the factorization*

$$\Omega(k) = L(k)\Lambda(k)L(k)'$$

with $L(k)$ and $\Lambda(k)$ described as in Theorem 10.1. Then we can write

$$L^{(-1)}(k) = L^{(-1,1)} + L^{(-1,0)}(k),$$

where $L^{(-1,1)}$ does not depend on k, and $L^{(-1,0)}(k)\Lambda^{(1)} = O(1/k)$.

Proof. Suppose first that $\Omega^{(1)}$ is diagonal with ith diagonal element $\omega_{ii}^{(1)}$. Then, as in the proof of Theorem 10.1, when $\omega_{jj}^{(1)} > 0$ we can write the ijth element of $L(k)$ as

$$l_{ij} = (1/k) l_{ij}^{(-1)} + O(1/k^2), \qquad i > j,$$

where $l_{ij}^{(-1)}$ does not depend on k. Define $L^{(-1,1)}$ to be the lower triangular matrix with ijth element $l_{ij}^{(-1,1)}$ given by

$$l_{ij}^{(-1,1)} = l_{ij}^{(-1)} \quad \text{if } i > j \quad \text{and} \quad \omega_{jj}^{(1)} > 0$$
$$= 0 \quad \text{otherwise.}$$

Define

$$L^{(-1,0)}(k) = L^{(-1)}(k) - L^{(-1,1)}.$$

Denoting the ijth element of $L^{(-1,0)}(k)$ by $l_{ij}^{(-1,0)}(k)$ we have

$$l_{ij}^{(-1,0)}(k) = O(1/k) \quad \text{if } \omega_{jj}^{(1)} > 0$$
$$= O(1) \quad \text{if } \omega_{jj}^{(1)} = 0.$$

Thus $L^{(-1,0)}\Lambda^{(1)} = O(1/k)$ as required.

To complete the proof note that we can again reduce the case of general $\Omega^{(1)}$ to the diagonal case as in the proof of Theorem 10.1. □

Let $\Omega(k)$ be a $q \times q$ matrix satisfying Condition I, so that by Theorem 10.1 we can factorize it as $L(k)\Lambda(k)L(k)'$. Now partition $L(k)$ and $\Lambda(k)$ as

$$L(k) = \begin{bmatrix} L_{11}(k) & 0 \\ L_{21}(k) & L_{22}(k) \end{bmatrix}, \qquad \Lambda(k) = \begin{bmatrix} \Lambda_{11}(k) & 0 \\ 0 & \Lambda_{22}(k) \end{bmatrix},$$

where $L_{11}(k)$ and $\Lambda_1(k)$ are $q_1 \times q_1$ matrices, and the other matrices in $L(k)$ and $\Lambda(k)$ are dimensioned conformally; $0 \le q_1 \le q$.

LEMMA 10.3. *Let the matrices $L(k)$ and $\Lambda(k)$ be partitioned as above, let $\Omega_{11}(k)$ be the upper left $q_1 \times q_1$ submatrix of $\Omega(k)$, and let*

$$S_{2.1}(k) = L_{22}(k)\Lambda_{22}(k)L_{22}(k)'.$$

Then

(i) *We can write $S_{2.1}(k)$ as*

$$S_{2.1}(k) = kS_{2.1}^{(1)} + S_{2.1}^{(0)} + (1/k)S_{2.1}^{(-1)}(k),$$

where

(10.14)

$$S_{2.1}^{(1)} = L_{22}^{(0)}\Lambda_{22}^{(1)}L_{22}^{(0)\prime}, \qquad S_{2.1}^{(0)} = L_{22}^{(0)}\Lambda_{22}^{(0)}L_{22}^{(0)\prime} + L_{22}^{(0)}\Lambda_{22}^{(1)}L_{22}^{(-1,1)\prime} + L^{(-1,1)}\Lambda_{22}^{(1)}L_{22}^{(0)\prime},$$

and $S_{2.1}^{(-1)}(k)$ has elements that are bounded in absolute value.

(ii) *Write $\Omega_{11}(k) = k\Omega_{11}^{(1)} + \Omega_{11}^{(0)}$. Then*

$$\text{rank}(S_{2.1}^{(1)}) = \text{rank}(\Omega^{(1)}) - \text{rank}(\Omega_{11}^{(1)}) \qquad (10.15)$$

and, in particular if

$$\text{rank}(\Omega^{(1)}) = \text{rank}(\Omega_{11}^{(1)}),$$

then

$$S_{2.1}^{(1)} = 0.$$

PROOF. Part (i) follows from Theorem 10.1, Corollary 10.3, and some simple algebra. To obtain Part (ii), note that

$$\operatorname{rank}(\Omega^{(1)}) = \operatorname{rank}(\Lambda_{11}^{(1)}) + \operatorname{rank}(\Lambda_{22}^{(1)}),$$

and that $\Omega_{11}^{(1)} = L_{11}^{(0)}\Lambda_{11}^{(1)}L_{11}^{(0)\prime}$, so that $\operatorname{rank}(\Omega_{11}^{(1)}) = \operatorname{rank}(\Lambda_{11}^{(1)})$. (10.15) follows. □

Many of the matrices we deal with in this paper are conditional covariance matrices and they satisfy the following condition.

CONDITION II. A matrix $S(k)$ satisfies condition II if it is constructed in the same way as the matrix $S_{2.1}(k)$ of Lemma 10.3.

We now obtain the following basic theorem, which also motivated Lemma 10.3.

THEOREM 10.2. *Let z be a zero mean q dimensional normal vector random variable, having variance–covariance matrix $\Omega(k)$ satisfying Condition I. Let z_1 consist of the first q_1 elements of z, and z_2 consist of the last q_2 elements of z, with $q_1 + q_2 = q$. Then,*

(i) $\operatorname{Var}(z_2|z_1) = S_{2.1}(k)$, *where $S_{2.1}(k)$ is defined in Lemma* 10.3.

(ii) *Let* $\Omega_{11}(k) = \operatorname{Var}(z_1)$. *Then* $\operatorname{rank}(S_{2.1}^{(1)}) = \operatorname{rank}(\Omega^{(1)}) - \operatorname{rank}(\Omega_{11}^{(1)})$. *In particular* $S_{2.1}^{(1)} = 0$ *if* $\operatorname{rank}(\Omega^{(1)}) = \operatorname{rank}(\Omega_{11}^{(1)})$.

(iii) *Suppose that we know a priori that* $\operatorname{rank}(\Omega^{(1)}) \le m$ *for some* $m \le q$. *If* $\operatorname{rank}(\Omega_{11}^{(1)}) = m$, *then* $S_{2.1}^{(1)} = 0$.

PROOF. Put $\varepsilon = L^{-1}(k)z$, and let ε_1 and ε_2 be the first q_1 and last q_2 elements of ε, respectively. Then the elements of ε are independent with variance–covariance matrix $\Lambda(k)$, and

$$z_1 = L_{11}\varepsilon_1, \qquad z_2 = L_{21}\varepsilon_1 + L_{22}\varepsilon_2$$

so that

$$\operatorname{Var}(z_2|z_1) = \operatorname{Var}(z_2|\varepsilon_1) = \operatorname{Var}(L_{22}\varepsilon_2) = L_{22}\Lambda_{22}L_{22}' = S_{2.1}(k).$$

This proves Part (i). Part (ii) is just Lemma 10.3(ii), and Part (iii) follows from Part (ii). □

We now repartition the submatrices (L_{21}, L_{22}) and Λ_{22} above as

$$(L_{21}(k), L_{22}(k)) = \begin{bmatrix} L_{31}(k) & L_{33}(k) & 0 \\ L_{41}(k) & L_{43}(k) & L_{44}(k) \end{bmatrix}$$

and

$$\Lambda_{22}(k) = \begin{bmatrix} \Lambda_{33}(k) & 0 \\ 0 & \Lambda_{44}(k) \end{bmatrix},$$

where L_{31} is $q_3 \times q_1$, L_{33} is $q_3 \times q_3$, L_{44} is $q_4 \times q_4$, Λ_{33} is $q_3 \times q_3$, and Λ_{44} is $q_4 \times q_4$. The other submatrices are dimensioned conformally, and $q_2 = q_3 + q_4$.

Now define the $q_3 \times q_3$ diagonal matrix $\tilde{\Lambda}$ as having jth diagonal element 1 if the jth diagonal element of Λ_{33} is not identically zero, and zero otherwise.

Let z_3 and z_4 consist of the first q_3 and the last q_4 elements, respectively, of z_2, where the vector z_2 is defined above.

THEOREM 10.3. *Let the vector random variables z, z_1, z_2, z_3, and z_4 be defined as above. Then*

$$E[\{z_4 - E(z_4|z_1)\}|\{z_3 - E(z_3|z_1)\}] = L_{43}^{(0)}\tilde{\Lambda}_{33}L_{33}^{(0)-1}\{z_3 - E(z_3|z_1)\} + O(1/k),$$

where $\tilde{\Lambda}_{33}$ is a diagonal matrix with ith *diagonal element* 1 *if the* ith *diagonal element of Λ_{33} is positive and zero otherwise.*

PROOF. From Theorem 10.1(i),

$$\mathrm{Var}([z_3, z_4]'|z_1) = \mathrm{Var}(z_2|z_1) = S_{2.1}(k),$$

where $S_{2.1}(k)$ is defined in Lemma 10.3 above, and

$$S_{2.1}(k) = \begin{bmatrix} L_{33}(k) & 0 \\ L_{43}(k) & LL_{44}(k) \end{bmatrix} \begin{bmatrix} \Lambda_{33}(k) & 0 \\ 0 & \Lambda_{44}(k) \end{bmatrix} \begin{bmatrix} L_{33}(k)' & L_{43}(k)' \\ 0 & L_{44}(k)' \end{bmatrix}.$$

Thus,

$$\mathrm{Cov}[(z_4 - E(z_4|z_1))(z_3 - E(z_3|z_1))] = L_{43}\Lambda_{33}L'_{33},$$
$$\mathrm{Var}[z_3|z_1] = L_{33}\Lambda_{33}L'_{33},$$

and therefore

$$\begin{aligned} E[(z_4 - E(z_4|z_1))|(z_3 - E(z_3|z_1))] &= (L_{43}\Lambda_{33}L'_{33})(L_{33}\Lambda_{33}L'_{33})^{-}(z_3 - E(z_3|z_1)) \\ &= L_{43}\tilde{\Lambda}_{33}L_{33}^{-1}(z_3 - E(z_3|z_1)) \\ &= L_{43}^{(0)}\tilde{\Lambda}_{33}L_{33}^{(0)-1}(z_3 - E(z_3|z_1)) + O(1/k), \end{aligned}$$

where $(L_{33}\Lambda_{33}L'_{33})^{-}$ denotes any pseudoinverse of $L_{33}\Lambda_{33}L'_{33}$.

10.2. Proof of results in Section 6.

PROOF OF LEMMA 6.1. From (2.4), (2.5), (4.3), and Assumption 5.1, any collection of $y(t_j)$, $y(t)$, and $x(t)$ has a variance–covariance matrix satisfying Condition I above. (6.2) is obtained from Theorem 10.3 if we take z_1 as the null vector and identify $x(t)$ with z_4 and $(y(t_1)', \dots, y(t_j)')'$ with z_3. (6.4) is obtained similarly. (6.3) is obtained from Theorem 10.2(i) and Lemma 10.3(i) if we identify $x(t)$ with z_2 and $(y(t_1)', \dots, y(t_j)')'$ with z_1. (6.5) is obtained similarly. □

PROOF OF LEMMA 6.2.

(10.16)

$$y(t+1) - y(t+1|t;k) = H(t+1)(x(t+1) - x(t+1|t;k)) + v(t+1)$$

and (6.6) follows. (6.7) follows from (6.6) and Lemma 6.1.

Let l be the largest integer such that $t_l \le t$. In Theorem 10.2(i) identify z_1 with $(y(t_1)', \dots, y(t_l)')'$ and z_2 with $(y(t+1)', x(t+1)')'$. Then (ii) follows from Theorem 10.2(i) and Corollary 10.3.

(6.8) and (6.9) are obtained similarly to the way (10.16) is obtained in Lemma 10.3(i).

If we identify z_4 with $x(t+1)$ and z_3 with $y(t+1)$, then (iv) follows from Theorem 10.3.

To obtain (v) note that from (10.16)

$$S_y^{(1)}(t+1|t) = H(t+1)S_x^{(1)}(t+1|t)H(t+1)', \tag{10.17}$$

so that from (6.6), (10.16), and (10.17)

$$S_{[y,x]}^{(1)}(t+1|t) = \begin{bmatrix} H(t+1) \\ I \end{bmatrix} S_x^{(1)}(t+1|t)[H(t+1)', I].$$

Therefore,

$$\begin{aligned} \operatorname{rank} S_{[y,x]}^{(1)}(t+1|t) &= \operatorname{rank} S_x^{(1)}(t+1|t) \\ &= \operatorname{rank} \Lambda_{yy}^{(1)} + \operatorname{rank} \Lambda_{xx}^{(1)} \\ &= \operatorname{rank} S_y^{(1)}(t+1|t) + \operatorname{rank} S_x^{(1)}(t+1|t+1) \end{aligned}$$

and this gives (v). □

PROOF OF THEOREM 6.1. (6.10) follows from (6.1). By the ordinary Kalman filter (Anderson and Moore, Chapter 3), $x(t+1|t) = F(t)x(t|t)$ and $S_x(t+1|t) = F(t)S_x(t|t)F(t)' + R(t)$ and (6.11) and (6.12) follow from Lemma 6.1. Step 2 is immediate.

(6.14) follows from Lemma 6.2(iv), and (6.15) from Lemma 6.2(v). □

10.3. Proof of Theorem 7.1. Part (i) follows from Lemma 6.2(i) except that we need to show that $\Lambda_t^{(1)}$ is independent of θ. From (2.5) and Assumption 2.5, $\operatorname{Var}(y) = kAA' + \Omega(\theta)$ where A is independent of θ and $\Omega(\theta)$ is nonsingular. Therefore, by Theorem 10.1 we can factorize $\operatorname{Var}(y)$ as $\tilde{L}(\theta; k)\tilde{\Lambda}(\theta; k)\tilde{L}(\theta; k)'$, with $\tilde{L}(\theta; k) = \tilde{L}^{(0)}(\theta) + O(1/k)$ and $\tilde{\Lambda}(\theta; k) = k\tilde{\Lambda}^{(1)} + \tilde{\Lambda}^{(0)} + O(1/k)$. $\tilde{L}$ is a lower triangular matrix having ones on the diagonal and $\tilde{\Lambda}$ is a diagonal matrix. By Theorem 10.1(v) $\tilde{\Lambda}^{(1)}$ is independent of θ because A is independent of θ.

Let $\tilde{L}_{jj}(\theta; k)$ and $\tilde{\Lambda}_{jj}(\theta; k)$ be the jjth $p \times p$ block submatrices of $\tilde{L}(\theta; k)$ and $\tilde{\Lambda}(\theta; k)$, respectively. Then from Theorem 10.2(i) and Lemma 10.3,

$$S_y(t_j|t_j - 1; \theta; k) = \tilde{L}_{jj}(\theta; k)\tilde{\Lambda}_{jj}(\theta; k)\tilde{L}_{jj}(\theta; k),$$

so that for $t = t_j$

$$\tilde{L}_{jj}^{(0)}(\theta)\tilde{\Lambda}_{jj}^{(1)}\tilde{L}_{jj}^{(0)}(\theta)' = L_j^{(0)}(\theta)\Lambda_j^{(1)}(\theta)L_j^{(0)}(\theta)',$$

implying that $\Lambda_j^{(1)}(\theta) = \tilde{\Lambda}_{jj}^{(1)}$ and hence $\Lambda_j^{(1)}$ is independent of θ.

As in Schweppe (1965), $-2 \times \log k^{D'/2} f(y;\theta;k)$ is given, up to an additive constant that does not depend on the parameters, by

$$\sum_{j=1}^{n} \log\det S_y(t_j|t_j-1) + \varepsilon(j)' S_y(t_j|t_j-1)^{-1}\varepsilon(j) - D'\log k$$

$$= \sum_{j=1}^{n}\sum_{i=1}^{p} \log\lambda_{ij} + \sum_{j=1}^{n}\sum_{i=1}^{p} \left(\xi_{ij} + O(1/k)\right)^2/\lambda_{ij} - D'\log k,$$

and as $k \to \infty$ this tends to

$$\sum_{j=1}^{n}\sum_{i\notin K_{j0}} \log\lambda_{ij}^{(0)} + \sum_{j=1}^{n}\sum_{i\in K_{j0}} \log\lambda_{ij}^{(0)} + \sum_{j=1}^{n}\sum_{i\in K_{j0}} \xi_{ij}^2/\lambda_{ij}^{(0)}.$$

Because $\lambda_{ij}^{(1)}$ is independent of θ, (7.1) follows. □

10.4. Proof of results in Section 8.

PROOF OF LEMMA 8.1.

$$x(t+1) - x(t+1|t) = F(t)(x(t) - x(t|t)) + u(t)$$

and (8.1) follows. Part (ii) is obtained similarly to Lemma 6.2(ii), and (8.2) and (8.3) are obtained as in the proof of Theorem 10.3(ii).

$$L_{11}^{-1}(k) = \left\{I - (1/k)\left(L_{11}^{(0)}\right)^{-1}\left[L_{11}^{(-1,1)} + L_{11}^{(-1,0)}(k)\right] + O(1/k^2)\right\}\left(L_{11}^{(0)}\right)^{-1},$$

and (8.4) to (8.6) can be obtained by simple algebra.

We now show that (8.7) holds. $S_x^{(1)}(t+1|t)$ is equal to $L_{11}^{(0)}\Lambda_{11}^{(1)}L_{11}^{(0)\prime}$ so that

$$C^{(-1,0)}(t;k)S_x^{(1)}(t+1|t) = \left[L_{21}^{(-1,0)}(k)\Lambda_{11}^{(1)} - L_{21}^{(0)}\tilde{\Lambda}_{11}\left(L_{11}^{(0)}\right)^{-1}L_{11}^{(-1,0)}(k)\Lambda_{11}^{(1)}\right]L_{11}^{(0)\prime}$$
$$= O(1/k)$$

because $L_{21}^{(-1,0)}(k)\Lambda_{11}^{(1)} = O(1/k) = L_{11}^{(-1,0)}(k)\Lambda_{11}^{(1)}$ by Lemma 8.1(ii). □

Before proving Theorem 8.2 we need the following lemmas.

LEMMA 10.4. *Suppose $\Omega(k)$ is a $q \times q$ positive semidefinite matrix for all positive k. Then $\Omega(k) = O(1/k^2)$ if and only if $\alpha'\Omega(k)\alpha = O(1/k^2)$ for all fixed $q \times 1$ vectors α.*

PROOF. That $\Omega(k) = O(1/k^2)$ implies $\alpha'\Omega(k)\alpha = O(1/k^2)$ is obvious. For the converse it is sufficient to consider the 2×2 case and take in turn $\alpha' = (1,0)$, $(0,1)$, and $(1,1)$. □

LEMMA 10.5. *Suppose that S is a $q \times q$ positive semidefinite matrix, and $T(k)$ a $p \times q$ matrix which depends on k. Then $T(k)S = O(1/k)$ if and only if $T(k)ST(k)' = O(1/k^2)$.*

PROOF. Assume first that S is a diagonal matrix with ith diagonal element s_{ii}. Let $T_i(k)$ be the ith column of T. Then $T(k)S = O(1/k)$ implies that $T_i(k)s_{ii}^{1/2} = O(1/k)$ for all i, so that

$$T(k)ST(k)' = \Sigma_i T_i(k)s_{ii}T_i(k)' = O(1/k^2). \tag{10.18}$$

Conversely, if (10.18) holds, then for each i $\mathrm{tr}\{T_i(k)s_{ii}T_i(k)'\} = O(1/k^2)$ so that $T_i(k)s_{ii}^{1/2} = O(1/k)$. This proves the result for S diagonal. The result holds for general S, because by Lemma 10.1 we can write $S = BDB'$ with B a lower triangular matrix having 1's on the diagonal and D a diagonal matrix. □

LEMMA 10.6. *Suppose that* (i) S_1, S_2, *and* $S_1 - S_2$ *are* $q \times q$ *positive semi-definite matrices.*

(ii) $T(k)$ *is a* $p \times q$ *matrix which depends on* k, *with elements that are uniformly bounded in absolute value.*

(iii) $T(k)S_1 = O(1/k)$.

Then $T(k)(S_1 - S_2) = O(1/k)$.

PROOF. By Lemma 10.5, $T(k)S_1T(k)' = O(1/k^2)$, so that for all α, $\alpha' T(k)S_1T(k)'\alpha = O(1/k^2)$ by Lemma 10.4. Therefore for all α

$$0 \le \alpha' T(k)(S_1 - S_2)T(k)'\alpha \le \alpha' T(k)S_1T(k)'\alpha = O(1/k^2),$$

and the required result now follows by applying Lemma 10.4 followed by Lemma 10.5. □

PROOF OF THEOREM 8.2. The ordinary fixed interval smoothing algorithm [Anderson and Moore (1979), Section 7.4] gives

$$x(t|N) = x(t|t) - C(t)(x(t+1|t) - x(t+1|N)) \tag{10.19}$$

and

$$S_x(t|N) = S_x(t|t) - C(t)(S_x(t+1|t) - S_x(t+1|N))C(t)' \tag{10.20}$$

with $C(t) = S_x(t|t)F(t)'S_x(t+1|t)^-$ and $S_x(t+1|t)^-$ is any pseudoinverse of $S_x(t+1|t)$. An alternative expression for $C(t)$ is given by (8.3).

(8.8) follows from (10.19) and Lemma 8.1(iii). From (8.7) and Lemma 10.6

$$C^{(-1,0)}(t;k)\big(S_x^{(1)}(t+1|t) - S_x^{(1)}(t+1|N)\big) = O(1/k)$$

and (8.9) and (8.10) now follow from (10.20) by some simple algebra. □

REFERENCES

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

ANSLEY, C. F. and KOHN, R. (1985). A structured state space approach to computing the likelihood of an ARIMA process and its derivatives. *J. Statist. Comput. Simulation* **21** 135–169.

BOX, G. E. P. and JENKINS, G. M. (1976). *Time Series Analysis, Forecasting and Control*, 2nd ed. Holden-Day, San Francisco.

HARVEY, A. C. and PHILLIPS, G. D. A. (1979). Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. *Biometrika* **66** 49–58.

HOCHSTADT, H. (1964). *Differential Equations*. Holt, Rinehart and Winston, New York.

JONES, R. H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics* **22** 389–395.

JONES, R. H. (1983). Fitting multivariate models to unequally spaced data. In *Time Series Analysis of Irregularly Observed Data* (E. Parzen, ed.). *Lecture Notes in Statistics* **25** 158–188. Springer, New York.

KOHN, R. and ANSLEY, C. F. (1983a). Fixed interval estimation in state space models when some of the data are missing or aggregated. *Biometrika* **70** 683–688.

KOHN, R. and ANSLEY, C. F. (1983b). Exact likelihood of a vector autoregressive-moving average process with missing or aggregated data. *Biometrika* **70** 275–276.

KOHN, R. and ANSLEY, C. F. (1984a). Estimation, prediction and interpolation for ARIMA models with missing data. Technical Report 24, Statistics Research Center, University of Chicago.

KOHN, R. and ANSLEY, C. F. (1984b). A new algorithm for spline smoothing and interpolation based on smoothing a stochastic process. Technical Report 21, Statistics Research Center, University of Chicago.

MENDEL, J. M. (1971). Computational requirements for a discrete Kalman filter. *IEEE Trans. Automat. Control* **AC-16**, 748–758.

SCHWEPPE, F. C. (1965). Evaluation of likelihood functions for Gaussian signals. *IEEE Trans. Inform. Theory* **IT-11** 61–70.

SCHWEPPE, F. C. (1973). *Uncertain Dynamic Systems*. Prentice-Hall, Englewood Cliffs.

WAHBA, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. Roy. Statist. Soc. Ser. B* **40** 364–372.

WEINERT, H. L., BYRD, R. H. and SIDHU, G. S. (1980). A stochastic framework for recursive computation of spline functions: Part II, smoothing splines. *J. Optim. Theory Appl.* **30** 255–268.

GRADUATE SCHOOL OF BUSINESS
UNIVERSITY OF CHICAGO
1101 E. 58TH STREET
CHICAGO, ILLINOIS 60637