5 Initialisation of filter and smoother

5.1 Introduction

In the previous chapter we have considered the operations of filtering and smoothing for the linear Gaussian state space model

$$
\begin{aligned}
y_t &= Z_t\alpha_t + \varepsilon_t, \qquad & \varepsilon_t &\sim \mathrm{N}(0, H_t), \\
\alpha_{t+1} &= T_t\alpha_t + R_t\eta_t, \qquad & \eta_t &\sim \mathrm{N}(0, Q_t),
\end{aligned}
\tag{5.1}
$$

under the assumption that $\alpha_1 \sim \mathrm{N}(a_1, P_1)$ where a_1 and P_1 are known. In most practical applications, however, at least some of the elements of a_1 and P_1 are unknown. We now develop methods of starting up the series when this is the situation; the process is called *initialisation*. We shall consider the general case where some elements of α_1 have a known joint distribution while about other elements we are completely ignorant. We treat the case in detail where the observations are normally distributed from a classical point of view. The results can be extended to minimum variance unbiased linear estimates and to Bayesian analysis by Lemmas 2, 3 and 4.

A general model for the initial state vector α_1 is

$$
\alpha_1 = a + A\delta + R_0\eta_0, \qquad \eta_0 \sim \mathrm{N}(0, Q_0),
\tag{5.2}
$$

where the $m \times 1$ vector a is known, δ is a $q \times 1$ vector of unknown quantities, the $m \times q$ matrix A and the $m \times (m-q)$ matrix R_0 are selection matrices, that is, they consist of columns of the identity matrix I_m; they are defined so that when taken together, their columns constitute a set of g columns of I_m with $g \leq m$ and $A'R_0 = 0$. The matrix Q_0 is assumed to be positive definite and known. In most cases vector a will be treated as a zero vector unless some elements of the initial state vector are known constants. When all elements of the state vector α_t are stationary, the initial means, variances and covariances of these initial state elements can be derived from the model parameters. For example, in the case of a stationary ARMA model it is straightforward to obtain the unconditional variance matrix Q_0 as we will show in Subsection 5.6.2. The Kalman filter (4.24) can then be applied routinely with $a_1 = 0$ and $P_1 = Q_0$.

To illustrate the structure and notation of (5.2) for readers unfamiliar with the subject, we present a simple example in which

$$
y_t = \mu_t + \rho_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{N}\big(0, \sigma_\varepsilon^2\big),
$$

where

$$\begin{aligned}
\mu_{t+1} &= \mu_t + \nu_t + \xi_t, & \xi_t &\sim \mathrm{N}\big(0, \sigma_\xi^2\big),\\
\nu_{t+1} &= \nu_t + \zeta_t, & \zeta_t &\sim \mathrm{N}\big(0, \sigma_\zeta^2\big),\\
\rho_{t+1} &= \phi\rho_t + \tau_t, & \tau_t &\sim \mathrm{N}\big(0, \sigma_\tau^2\big),
\end{aligned}$$

in which $|\phi| < 1$ and the disturbances are all mutually and serially uncorrelated. Thus μ_t is a local linear trend as in (3.2), which is nonstationary, while ρ_t is an unobserved stationary first order AR(1) series with zero mean. In state space form this is

$$y_t = (1 \quad 0 \quad 1)\begin{pmatrix}\mu_t\\ \nu_t\\ \rho_t\end{pmatrix} + \varepsilon_t,$$

$$\begin{pmatrix}\mu_{t+1}\\ \nu_{t+1}\\ \rho_{t+1}\end{pmatrix} = \begin{bmatrix}1 & 1 & 0\\ 0 & 1 & 0\\ 0 & 0 & \phi\end{bmatrix}\begin{pmatrix}\mu_t\\ \nu_t\\ \rho_t\end{pmatrix} + \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{pmatrix}\xi_t\\ \zeta_t\\ \tau_t\end{pmatrix}.$$

Thus we have

$$a = \begin{pmatrix}0\\0\\0\end{pmatrix}, \qquad A = \begin{bmatrix}1 & 0\\ 0 & 1\\ 0 & 0\end{bmatrix}, \qquad R_0 = \begin{pmatrix}0\\0\\1\end{pmatrix},$$

with $\eta_0 = \rho_1$ and where $Q_0 = \sigma_\tau^2/(1-\phi^2)$ is the variance of the stationary series ρ_t.

Although we treat the parameter ϕ as known for the purpose of this section, in practice it is unknown, which in a classical analysis is replaced by its maximum likelihood estimate. We see that the object of the representation (5.2) is to separate out α_1 into a constant part a, a nonstationary part $A\delta$ and a stationary part $R_0\eta_0$. In a Bayesian analysis, α_1 can be treated as having a known or noninformative prior density.

The vector δ can be treated as a fixed vector of unknown parameters or as a vector of random normal variables with infinite variances. For the case where δ is fixed and unknown, we may estimate it by maximum likelihood; a technique for doing this was developed by Rosenberg (1973) and we will discuss this in Section 5.7. For the case where δ is random we assume that

$$\delta \sim \mathrm{N}(0, \kappa I_q), \tag{5.3}$$

where we let $\kappa \to \infty$. We begin by considering the Kalman filter with initial conditions $a_1 = \mathrm{E}(\alpha_1) = a$ and $P_1 = \mathrm{Var}(\alpha_1)$ where

$$P_1 = \kappa P_\infty + P_*, \tag{5.4}$$

and we let $\kappa \to \infty$ at a suitable point later. Here $P_\infty = AA'$ and $P_* = R_0 Q_0 R_0'$; since A consists of columns of I_m it follows that P_∞ is an $m \times m$ diagonal matrix with q diagonal elements equal to one and the other elements equal to zero. Also, without loss of generality, when a diagonal element of P_∞ is nonzero we take the corresponding element of a to be zero. A vector δ with distribution $\mathrm{N}(0, \kappa I_q)$ as $\kappa \to \infty$ is said to be *diffuse*. Initialisation of the Kalman filter when some elements of α_1 are diffuse is called *diffuse initialisation* of the filter. We now consider the modifications required to the Kalman filter in the diffuse initialisation case.

A simple approximate technique is to replace κ in (5.4) by an arbitrary large number and then use the standard Kalman filter (4.13). This approach was employed by Harvey and Phillips (1979). While the device can be useful for approximate exploratory work, it is not recommended for general use since it can lead to large rounding errors. We therefore develop an exact treatment.

The technique we shall use is to expand matrix products as power series in κ^{-1}, taking only the first two or three terms as required, and then let $\kappa \to \infty$ to obtain the dominant term. The underlying idea was introduced by Ansley and Kohn (1985) in a somewhat inaccessible way. Koopman (1997) presented a more transparent treatment of diffuse filtering and smoothing based on the same idea. Further developments were given by Koopman and Durbin (2003) who obtained the results that form the basis of Section 5.2 on filtering and Section 5.3 on state smoothing. This approach gives recursions different from those obtained from the augmentation technique of de Jong (1991) which is based on ideas of Rosenberg (1973); see Section 5.7. Illustrations of these initialisation methods are given in Section 5.6 and Subsection 5.7.4.

A direct approach to the initialisation problem for the general multivariate linear Gaussian state space model turns out to be somewhat complicated as can be seen from the treatment of Koopman (1997). The reason for this is that for multivariate series the inverse matrix F_t^{-1} does not have a simple general expansion in powers of κ^{-1} for the first few terms of the series, due to the fact that in very specific situations the part of F_t associated with P_∞ can be singular with varying rank. Rank deficiencies may occur when observations are missing near the beginning of the series, for example. For univariate series, however, the treatment is much simpler since F_t is a scalar so the part associated with P_∞ can only be either zero or positive, both of which are easily dealt with. In complicated cases, it turns out to be simpler in the multivariate case to adopt the filtering and smoothing approach of Section 6.4 in which the multivariate series is converted to a univariate series by introducing the elements of the observational vector y_t one at a time, rather than dealing with the series directly as a multivariate series. We therefore begin by assuming that the part of F_t associated with P_∞ is nonsingular or zero for any t. In this way we can treat most multivariate series, for which this assumption holds directly, and at the same time obtain general results for all univariate time series. We shall use these results in Section 6.4 for the univariate treatment of multivariate series.

5.2 The exact initial Kalman filter

In this section we use the notation $O(\kappa^{-j})$ to denote a function $f(\kappa)$ of κ such that the limit of $\kappa^j f(\kappa)$ as $\kappa \to \infty$ is finite for $j = 1, 2$.

5.2.1 The basic recursions

Analogously to the decomposition of the initial matrix P_1 in (5.4) we show that the mean square error matrix P_t has the decomposition

$$P_t = \kappa P_{\infty,t} + P_{*,t} + O(\kappa^{-1}), \qquad t = 2, \dots, n, \tag{5.5}$$

where $P_{\infty,t}$ and $P_{*,t}$ do not depend on κ. It will be shown that $P_{\infty,t} = 0$ for $t > d$ where d is a positive integer which in normal circumstances is small relative to n. The consequence is that the usual Kalman filter (4.24) applies without change for $t = d+1, \dots, n$ with $P_t = P_{*,t}$. Note that when all initial state elements have a known joint distribution or are fixed and known, matrix $P_\infty = 0$ and therefore $d = 0$.

The decomposition (5.5) leads to the similar decompositions

$$F_t = \kappa F_{\infty,t} + F_{*,t} + O(\kappa^{-1}), \qquad M_t = \kappa M_{\infty,t} + M_{*,t} + O(\kappa^{-1}), \tag{5.6}$$

and, since $F_t = Z_t P_t Z_t' + H_t$ and $M_t = P_t Z_t'$, we have

$$\begin{aligned} F_{\infty,t} &= Z_t P_{\infty,t} Z_t', & F_{*,t} &= Z_t P_{*,t} Z_t' + H_t, \\ M_{\infty,t} &= P_{\infty,t} Z_t', & M_{*,t} &= P_{*,t} Z_t', \end{aligned} \tag{5.7}$$

for $t = 1, \dots, d$. The Kalman filter that we shall derive as $\kappa \to \infty$ we shall call the *exact initial Kalman filter*. We use the word *exact* here to distinguish the resulting filter from the approximate filter obtained by choosing an arbitrary large value for κ and applying the standard Kalman filter (4.24). In deriving it, it is important to note from (5.7) that a zero matrix $M_{\infty,t}$ (whether $P_{\infty,t}$ is a zero matrix or not) implies that $F_{\infty,t} = 0$. As in the development of the Kalman filter in Subsection 4.3.2 we assume that F_t is nonsingular. The derivation of the exact initial Kalman filter is based on the expansion for $F_t^{-1} = [\kappa F_{\infty,t} + F_{*,t} + O(\kappa^{-1})]^{-1}$ as a power series in κ^{-1}, that is

$$F_t^{-1} = F_t^{(0)} + \kappa^{-1} F_t^{(1)} + \kappa^{-2} F_t^{(2)} + O(\kappa^{-3}), \tag{5.8}$$

for large κ. Since $I_p = F_t F_t^{-1}$ we have

$$\begin{aligned} I_p = {} & (\kappa F_{\infty,t} + F_{*,t} + \kappa^{-1} F_{a,t} + \kappa^{-2} F_{b,t} + \cdots) \\ & \times \left(F_t^{(0)} + \kappa^{-1} F_t^{(1)} + \kappa^{-2} F_t^{(2)} + \cdots\right). \end{aligned}$$

On equating coefficients of κ^j for $j = 0, -1, -2, \ldots$ we obtain

$$
\begin{aligned}
F_{\infty,t}F_t^{(0)} &= 0, \\
F_{*,t}F_t^{(0)} + F_{\infty,t}F_t^{(1)} &= I_p, \\
F_{a,t}F_t^{(0)} + F_{*,t}F_t^{(1)} + F_{\infty,t}F_t^{(2)} &= 0, \quad \text{etc.}
\end{aligned}
\tag{5.9}
$$

We need to solve equations (5.9) for $F_t^{(0)}$, $F_t^{(1)}$ and $F_t^{(2)}$; further terms are not required. We shall consider only the cases where $F_{\infty,t}$ is nonsingular or $F_{\infty,t} = 0$. This limitation of the treatment is justified for three reasons. First, it gives a complete solution for the important special case of univariate series, since if y_t is univariate $F_{\infty,t}$ must obviously be positive or zero. Second, if y_t is multivariate the restriction is satisfied in most practical cases. Third, for those rare cases where y_t is multivariate but the restriction is not satisfied, the series can be dealt with as a univariate series by the technique described in Section 6.4. By limiting the treatment at this point to these two cases, the derivations are essentially no more difficult than those required for treating the univariate case. However, solutions for the general case where no restrictions are placed on $F_{\infty,t}$ are algebraically complicated; see Koopman (1997). We mention that although $F_{\infty,t}$ nonsingular is the most common case, situations can arise in practice where $F_{\infty,t} = 0$ even when $P_{\infty,t} \neq 0$ if $M_{\infty,t} = P_{\infty,t}Z_t' = 0$.

Taking first the case where $F_{\infty,t}$ is nonsingular we have from (5.9),

$$
F_t^{(0)} = 0, \qquad F_t^{(1)} = F_{\infty,t}^{-1}, \qquad F_t^{(2)} = -F_{\infty,t}^{-1}F_{*,t}F_{\infty,t}^{-1}. \tag{5.10}
$$

The matrices $K_t = T_t M_t F_t^{-1}$ and $L_t = T_t - K_t Z_t$ depend on the inverse matrix F_t^{-1} so they also can be expressed as power series in κ^{-1}. We have

$$
K_t = T_t[\kappa M_{\infty,t} + M_{*,t} + O(\kappa^{-1})]\big(\kappa^{-1}F_t^{(1)} + \kappa^{-2}F_t^{(2)} + \cdots\big),
$$

so

$$
K_t = K_t^{(0)} + \kappa^{-1}K_t^{(1)} + O(\kappa^{-2}), \qquad L_t = L_t^{(0)} + \kappa^{-1}L_t^{(1)} + O(\kappa^{-2}), \tag{5.11}
$$

where

$$
\begin{aligned}
K_t^{(0)} &= T_t M_{\infty,t}F_t^{(1)}, & L_t^{(0)} &= T_t - K_t^{(0)}Z_t, \\
K_t^{(1)} &= T_t M_{*,t}F_t^{(1)} + T_t M_{\infty,t}F_t^{(2)}, & L_t^{(1)} &= -K_t^{(1)}Z_t.
\end{aligned}
\tag{5.12}
$$

By following the recursion (4.21) for a_{t+1} starting with $t = 1$ we find that a_t has the form

$$
a_t = a_t^{(0)} + \kappa^{-1}a_t^{(1)} + O(\kappa^{-2}),
$$

where $a_1^{(0)} = a$ and $a_1^{(1)} = 0$. As a consequence v_t has the form

$$
v_t = v_t^{(0)} + \kappa^{-1}v_t^{(1)} + O(\kappa^{-2}),
$$

where $v_t^{(0)} = y_t - Z_t a_t^{(0)}$ and $v_t^{(1)} = -Z_t a_t^{(1)}$. The updating equation (4.21) for a_{t+1} can now be expressed as

$$\begin{aligned} a_{t+1} &= T_t a_t + K_t v_t \\ &= T_t\big[a_t^{(0)} + \kappa^{-1} a_t^{(1)} + O(\kappa^{-2})\big] \\ &\quad + \big[K_t^{(0)} + \kappa^{-1} K_t^{(1)} + O(\kappa^{-2})\big]\big[v_t^{(0)} + \kappa^{-1} v_t^{(1)} + O(\kappa^{-2})\big], \end{aligned}$$

which becomes as $\kappa \to \infty$,

$$a_{t+1}^{(0)} = T_t a_t^{(0)} + K_t^{(0)} v_t^{(0)}, \qquad t = 1, \dots, n. \tag{5.13}$$

The updating equation (4.23) for P_{t+1} is

$$\begin{aligned} P_{t+1} &= T_t P_t L_t' + R_t Q_t R_t' \\ &= T_t[\kappa P_{\infty,t} + P_{*,t} + O(\kappa^{-1})]\big[L_t^{(0)} + \kappa^{-1} L_t^{(1)} + O(\kappa^{-2})\big]' + R_t Q_t R_t'. \end{aligned}$$

Consequently, on letting $\kappa \to \infty$, the updates for $P_{\infty,t+1}$ and $P_{*,t+1}$ are given by

$$\begin{aligned} P_{\infty,t+1} &= T_t P_{\infty,t} L_t^{(0)\prime}, \\ P_{*,t+1} &= T_t P_{\infty,t} L_t^{(1)\prime} + T_t P_{*,t} L_t^{(0)\prime} + R_t Q_t R_t', \end{aligned} \tag{5.14}$$

for $t = 1, \dots, n$. The matrix P_{t+1} also depends on terms in κ^{-1}, κ^{-2}, etc. but these terms will not be multiplied by κ or higher powers of κ within the Kalman filter recursions. Thus the updating equations for P_{t+1} do not need to take account of these terms. Recursions (5.13) and (5.14) constitute the exact Kalman filter.

In the case where $F_{\infty,t} = 0$, we have

$$F_t = F_{*,t} + O(\kappa^{-1}), \qquad M_t = M_{*,t} + O(\kappa^{-1}),$$

and the inverse matrix F_t^{-1} is given by

$$F_t^{-1} = F_{*,t}^{-1} + O(\kappa^{-1}).$$

Therefore,

$$\begin{aligned} K_t &= T_t[M_{*,t} + O(\kappa^{-1})]\big[F_{*,t}^{-1} + O(\kappa^{-1})\big] \\ &= T_t M_{*,t} F_{*,t}^{-1} + O(\kappa^{-1}). \end{aligned}$$

The updating equation for $a_{t+1}^{(0)}$ (5.13) has

$$K_t^{(0)} = T_t M_{*,t} F_{*,t}^{-1}, \tag{5.15}$$

and the updating equation for P_{t+1} becomes

$$\begin{aligned} P_{t+1} &= T_t P_t L_t' + R_t Q_t R_t' \\ &= T_t[\kappa P_{\infty,t} + P_{*,t} + O(\kappa^{-1})]\big[L_t^{(0)} + \kappa^{-1} L_t^{(1)} + O(\kappa^{-2})\big]' + R_t Q_t R_t', \end{aligned}$$

where $L_t^{(0)} = T_t - K_t^{(0)} Z_t$ and $L_t^{(1)} = -K_t^{(1)} Z_t$. The updates for $P_{\infty,t+1}$ and $P_{*,t}$ can be simplified considerably since $M_{\infty,t} = P_{\infty,t} Z_t' = 0$ when $F_{\infty,t} = 0$. By letting $\kappa \to \infty$ we have

$$\begin{aligned} P_{\infty,t+1} &= T_t P_{\infty,t} L_t^{(0)\prime} \\ &= T_t P_{\infty,t} T_t' - T_t P_{\infty,t} Z_t' K_t^{(0)\prime} \\ &= T_t P_{\infty,t} T_t', \qquad (5.16) \\ P_{*,t+1} &= T_t P_{\infty,t} L_t^{(1)\prime} + T_t P_{*,t} L_t^{(0)\prime} + R_t Q R_t' \\ &= -T_t P_{\infty,t} Z_t' K_t^{(1)\prime} + T_t P_{*,t} L_t^{(0)\prime} + R_t Q R_t' \\ &= T_t P_{*,t} L_t^{(0)\prime} + R_t Q R_t', \qquad (5.17) \end{aligned}$$

for $t = 1, \ldots, d$, with $P_{\infty,1} = P_\infty = AA'$ and $P_{*,1} = P_* = R_0 Q_0 R_0'$. It might be thought that an expression of the form $F_{*,t} + \kappa^{-1} F_{**,t} + O(\kappa^{-2})$ should be used for F_t here so that two-term expansions could be carried out throughout. It can be shown however that when $M_{\infty,t} = P_{\infty,t} Z_t' = 0$, so that $F_{\infty,t} = 0$, the contribution of the term $\kappa^{-1} F_{**,t}$ is zero; we have therefore omitted it to simplify the presentation.

5.2.2 Transition to the usual Kalman filter

We now show that for nondegenerate models there is a value of d of t such that $P_{\infty,t} \neq 0$ for $t \leq d$ and $P_{\infty,t} = 0$ for $t > d$. From (5.2) the vector of diffuse elements of α_1 is δ and its dimensionality is q. For finite κ the logdensity of δ is

$$\log p(\delta) = -\frac{q}{2}\log 2\pi - \frac{q}{2}\log\kappa - \frac{1}{2\kappa}\delta'\delta,$$

since $\mathrm{E}(\delta) = 0$ and $\mathrm{Var}(\delta) = \kappa I_q$. Now consider the joint density of δ and Y_t. In an obvious notation the log conditional density of δ given Y_t is

$$\log p(\delta|Y_t) = \log p(\delta, Y_t) - \log p(Y_t),$$

for $t = 1, \ldots, n$. Differentiating with respect to δ, letting $\kappa \to \infty$, equating to zero and solving for δ, gives a solution $\delta = \tilde{\delta}$ which is the conditional mode, and hence the conditional mean, of δ given Y_t.

Since $p(\delta, Y_t)$ is Gaussian, $\log p(\delta, Y_t)$ is quadratic in δ so its second derivative does not depend on δ. The reciprocal of minus the second derivative is the variance matrix of δ given Y_t. Let d be the first value of t for which this variance matrix exists. In practical cases d will usually be small relative to n. If there is no value of t for which the variance matrix exists we say that the model is degenerate, since a series of observations which does not even contain enough information to estimate the initial conditions is clearly useless.

By repeated substitution from the state equation $\alpha_{t+1} = T_t\alpha_t + R_t\eta_t$ we can express α_{t+1} as a linear function of α_1 and $\eta_1, \dots, \eta_t$. Elements of α_1 other than those in δ and also elements of $\eta_1, \dots, \eta_t$ have finite unconditional variances and hence have finite conditional variances given Y_t. We have also ensured that elements of δ have finite conditional variances given Y_t for $t \geq d$ by definition of d. It follows that $\text{Var}(\alpha_{t+1}|Y_t) = P_{t+1}$ is finite and hence $P_{\infty,t+1} = 0$ for $t \geq d$. On the other hand, for $t < d$, elements of $\text{Var}(\delta|Y_t)$ become infinite as $\kappa \to \infty$ from which it follows that elements of $\text{Var}(\alpha_{t+1}|Y_t)$ become infinite, so $P_{\infty,t+1} \neq 0$ for $t < d$. This establishes that for nondegenerate models there is a value d of t such that $P_{\infty,t} \neq 0$ for $t \leq d$ and $P_{\infty,t} = 0$ for $t > d$. Thus when $t > d$ we have $P_t = P_{*,t} + O(\kappa^{-1})$ so on letting $\kappa \to \infty$ we can use the usual Kalman filter (4.24) starting with $a_{d+1} = a_{d+1}^{(0)}$ and $P_{d+1} = P_{*,d+1}$. A similar discussion of this point is given by de Jong (1991).

5.2.3 A convenient representation

The updating equations for $P_{*,t+1}$ and $P_{\infty,t+1}$, for $t = 1, \dots, d$, can be combined to obtain a very convenient representation. Let

$$P_t^\dagger = [P_{*,t} \quad P_{\infty,t}], \qquad L_t^\dagger = \begin{bmatrix} L_t^{(0)} & L_t^{(1)} \\ 0 & L_t^{(0)} \end{bmatrix}. \tag{5.18}$$

From (5.14), the limiting initial state filtering equations as $\kappa \to \infty$ can be written as

$$P_{t+1}^\dagger = T_t P_t^\dagger L_t^{\dagger\prime} + [R_t Q_t R_t' \quad 0], \qquad t = 1, \dots, d, \tag{5.19}$$

with the initialisation $P_1^\dagger = P^\dagger = [P_* \quad P_\infty]$. For the case $F_{\infty,t} = 0$, the equations in (5.19) with the definitions in (5.18) are still valid but with

$$K_t^{(0)} = T_t M_{*,t} F_{*,t}^{-1}, \qquad L_t^{(0)} = T_t - K_t^{(0)} Z_t, \qquad L_t^{(1)} = 0.$$

This follows directly from the argument used to derive (5.15), (5.16) and (5.17). The recursion (5.19) for diffuse state filtering is due to Koopman and Durbin (2003). It is similar in form to the standard Kalman filtering (4.24) recursion, leading to simplifications in implementing the computations.

5.3 Exact initial state smoothing

5.3.1 Smoothed mean of state vector

To obtain the limiting recursions for the smoothing equation $\hat{\alpha}_t = a_t + P_t r_{t-1}$ given in (4.39) for $t = d, \dots, 1$, we return to the recursion (4.38) for r_{t-1}, that is,

$$r_{t-1} = Z_t' F_t^{-1} v_t + L_t' r_t, \qquad t = n, \dots, 1,$$

with $r_n = 0$. Since r_{t-1} depends on F_t^{-1} and L_t which can both be expressed as power series in κ^{-1} we write

$$r_{t-1} = r_{t-1}^{(0)} + \kappa^{-1} r_{t-1}^{(1)} + O(\kappa^{-2}), \qquad t = d, \dots, 1. \tag{5.20}$$

Substituting the relevant expansions into the recursion for r_{t-1} we have for the case $F_{\infty,t}$ nonsingular,

$$\begin{aligned} r_{t-1}^{(0)} + \kappa^{-1} r_{t-1}^{(1)} + \cdots = {} & Z_t' \big(\kappa^{-1} F_t^{(1)} + \kappa^{-2} F_t^{(2)} + \cdots\big)\big(v_t^{(0)} + \kappa^{-1} v_t^{(1)} + \cdots\big) \\ & + \big(L_t^{(0)} + \kappa^{-1} L_t^{(1)} + \cdots\big)'\big(r_t^{(0)} + \kappa^{-1} r_t^{(1)} + \cdots\big), \end{aligned}$$

leading to recursions for $r_t^{(0)}$ and $r_t^{(1)}$,

$$\begin{aligned} r_{t-1}^{(0)} &= L_t^{(0)\prime} r_t^{(0)}, \\ r_{t-1}^{(1)} &= Z_t' F_t^{(1)} v_t^{(0)} + L_t^{(0)\prime} r_t^{(1)} + L_t^{(1)\prime} r_t^{(0)}, \end{aligned} \tag{5.21}$$

for $t = d, \dots, 1$ with $r_d^{(0)} = r_d$ and $r_d^{(1)} = 0$.

The smoothed state vector is

$$\begin{aligned} \hat{\alpha}_t &= a_t + P_t r_{t-1} \\ &= a_t + [\kappa P_{\infty,t} + P_{*,t} + O(\kappa^{-1})]\big[r_{t-1}^{(0)} + \kappa^{-1} r_{t-1}^{(1)} + O(\kappa^{-2})\big] \\ &= a_t + \kappa P_{\infty,t}\big(r_{t-1}^{(0)} + \kappa^{-1} r_{t-1}^{(1)}\big) + P_{*,t}\big(r_{t-1}^{(0)} + \kappa^{-1} r_{t-1}^{(1)}\big) + O(\kappa^{-1}) \\ &= a_t + \kappa P_{\infty,t} r_{t-1}^{(0)} + P_{*,t} r_{t-1}^{(0)} + P_{\infty,t} r_{t-1}^{(1)} + O(\kappa^{-1}), \end{aligned} \tag{5.22}$$

where $a_t = a_t^{(0)} + \kappa^{-1} a_t^{(1)} + \cdots$. It is immediately obvious that for this expression to make sense we must have $P_{\infty,t} r_{t-1}^{(0)} = 0$ for all t. This will be the case if we can show that $\mathrm{Var}(\alpha_t|Y_n)$ is finite for all t as $\kappa \to \infty$. Analogously to the argument in Subsection 5.2.2 we can express α_t as a linear function of $\delta, \eta_0, \eta_1, \dots, \eta_{t-1}$. But $\mathrm{Var}(\delta|Y_d)$ is finite by definition of d so $\mathrm{Var}(\delta|Y_n)$ must be finite as $\kappa \to \infty$ since $d < n$. Also, $Q_j = \mathrm{Var}(\eta_j)$ is finite so $\mathrm{Var}(\eta_j|Y_n)$ is finite for $j = 0, \dots, t-1$. It follows that $\mathrm{Var}(\alpha_t|Y_n)$ is finite for all t as $\kappa \to \infty$ so from (5.22) $P_{\infty,t} r_{t-1}^{(0)} = 0$.

Letting $\kappa \to \infty$ we obtain

$$\hat{\alpha}_t = a_t^{(0)} + P_{*,t} r_{t-1}^{(0)} + P_{\infty,t} r_{t-1}^{(1)}, \qquad t = d, \dots, 1, \tag{5.23}$$

with $r_d^{(0)} = r_d$ and $r_d^{(1)} = 0$. The equations (5.21) and (5.23) can be re-formulated to obtain

$$r_{t-1}^{\dagger} = \begin{pmatrix} 0 \\ Z_t' F_t^{(1)} v_t^{(0)} \end{pmatrix} + L_t^{\dagger\prime} r_t^{\dagger}, \qquad \hat{\alpha}_t = a_t^{(0)} + P_t^{\dagger} r_{t-1}^{\dagger}, \qquad t = d, \dots, 1, \tag{5.24}$$

where

$$r^{\dagger}_{t-1} = \begin{pmatrix} r^{(0)}_{t-1} \\ r^{(1)}_{t-1} \end{pmatrix}, \quad \text{with} \quad r^{\dagger}_{d} = \begin{pmatrix} r_d \\ 0 \end{pmatrix},$$

and the partioned matrices $P^{\dagger}_t$ and $L^{\dagger}_t$ are defined in (5.18). This formulation is convenient since it has the same form as the standard smoothing recursion (4.39). Considering the complexity introduced into the model by the presence of the diffuse elements in α_1, it is very interesting that the state smoothing equations in (5.24) have the same basic structure as the corresponding equations (4.39). This is a useful property in constructing software for implementation of the algorithms.

To avoid extending the treatment further, and since the case $F_{\infty,t} = 0$ is rare in practice, when $P_{\infty,t} \neq 0$, we omit consideration of it here and in Subsection 5.3.2 and refer the reader to the discussion in Koopman and Durbin (2003).

5.3.2 Smoothed variance of state vector

We now consider the evaluation of the variance matrix of the estimation error $\hat{\alpha}_t - \alpha_t$ for $t = d, \ldots, 1$ in the diffuse case. We shall not derive the cross-covariances between the estimation errors at different time points in the diffuse case because in practice there is little interest in these quantities.

From Subsection 4.4.3, the error variance matrix of the smoothed state vector is given by $V_t = P_t - P_t N_{t-1} P_t$ with the recursion $N_{t-1} = Z_t' F_t^{-1} Z_t + L_t' N_t L_t$, for $t = n, \ldots, 1$, and $N_n = 0$. To obtain exact finite expressions for V_t and N_{t-1} where $F_{\infty,t}$ is nonsingular and $\kappa \to \infty$, for $t = d, \ldots, 1$, we find that we need to take three-term expansions instead of the two-term expressions previously employed. Thus we write

$$N_t = N_t^{(0)} + \kappa^{-1} N_t^{(1)} + \kappa^{-2} N_t^{(2)} + O(\kappa^{-3}). \tag{5.25}$$

Ignoring residual terms and on substituting in the expression for N_{t-1}, we obtain the recursion for N_{t-1} as

$$\begin{aligned} N^{(0)}_{t-1} &+ \kappa^{-1} N^{(1)}_{t-1} + \kappa^{-2} N^{(2)}_{t-1} + \cdots \\ &= Z_t' \big(\kappa^{-1} F_t^{(1)} + \kappa^{-2} F_t^{(2)} + \cdots\big) Z_t + \big(L_t^{(0)} + \kappa^{-1} L_t^{(1)} + \kappa^{-2} L_t^{(2)} + \cdots\big)' \\ &\quad \times \big(N_t^{(0)} + \kappa^{-1} N_t^{(1)} + \kappa^{-2} N_t^{(2)} + \cdots\big)\big(L_t^{(0)} + \kappa^{-1} L_t^{(1)} + \kappa^{-2} L_t^{(2)} + \cdots\big), \end{aligned}$$

which leads to the set of recursions

$$
\begin{aligned}
N_{t-1}^{(0)} &= L_t^{(0)\prime} N_t^{(0)} L_t^{(0)},\\
N_{t-1}^{(1)} &= Z_t' F_t^{(1)} Z_t + L_t^{(0)\prime} N_t^{(1)} L_t^{(0)} + L_t^{(1)\prime} N_t^{(0)} L_t^{(0)} + L_t^{(0)\prime} N_t^{(0)} L_t^{(1)},\\
N_{t-1}^{(2)} &= Z_t' F_t^{(2)} Z_t + L_t^{(0)\prime} N_t^{(2)} L_t^{(0)} + L_t^{(0)\prime} N_t^{(1)} L_t^{(1)} + L_t^{(1)\prime} N_t^{(1)} L_t^{(0)}\\
&\quad + L_t^{(0)\prime} N_t^{(0)} L_t^{(2)} + L_t^{(2)\prime} N_t^{(0)} L_t^{(0)} + L_t^{(1)\prime} N_t^{(0)} L_t^{(1)},
\end{aligned} \tag{5.26}
$$

with $N_d^{(0)} = N_d$ and $N_d^{(1)} = N_d^{(2)} = 0$.

Substituting the power series in κ^{-1}, κ^{-2}, etc. and the expression $P_t = \kappa P_{\infty,t} + P_{*,t}$ into the relation $V_t = P_t - P_t N_{t-1} P_t$ we obtain

$$
\begin{aligned}
V_t &= \kappa P_{\infty,t} + P_{*,t}\\
&\qquad - (\kappa P_{\infty,t} + P_{*,t})\big(N_{t-1}^{(0)} + \kappa^{-1} N_{t-1}^{(1)} + \kappa^{-2} N_{t-1}^{(2)} + \cdots\big)(\kappa P_{\infty,t} + P_{*,t})\\
&= -\kappa^2 P_{\infty,t} N_{t-1}^{(0)} P_{\infty,t}\\
&\qquad + \kappa\big(P_{\infty,t} - P_{\infty,t} N_{t-1}^{(0)} P_{*,t} - P_{*,t} N_{t-1}^{(0)} P_{\infty,t} - P_{\infty,t} N_{t-1}^{(1)} P_{\infty,t}\big)\\
&\qquad + P_{*,t} - P_{*,t} N_{t-1}^{(0)} P_{*,t} - P_{*,t} N_{t-1}^{(1)} P_{\infty,t} - P_{\infty,t} N_{t-1}^{(1)} P_{*,t}\\
&\qquad - P_{\infty,t} N_{t-1}^{(2)} P_{\infty,t} + O(\kappa^{-1}).
\end{aligned} \tag{5.27}
$$

It was shown in the previous section that $V_t = \mathrm{Var}(\alpha_t | Y_n)$ is finite for $t = 1, \ldots, n$. Thus the two matrix terms associated with κ and κ^2 in (5.27) must be zero. Letting $\kappa \to \infty$, the smoothed state variance matrix is given by

$$
V_t = P_{*,t} - P_{*,t} N_{t-1}^{(0)} P_{*,t} - P_{*,t} N_{t-1}^{(1)} P_{\infty,t} - P_{\infty,t} N_{t-1}^{(1)} P_{*,t} - P_{\infty,t} N_{t-1}^{(2)} P_{\infty,t}. \tag{5.28}
$$

Koopman and Durbin (2003) have shown by exploiting the equality $P_{\infty,t} L_t^{(0)} N_t^{(0)} = 0$, that when the recursions for $N_t^{(1)}$ and $N_t^{(2)}$ in (5.26) are used to calculate the terms in $N_{t-1}^{(1)}$ and $N_{t-1}^{(2)}$, respectively, in (5.28), various items vanish and that the effect is that we can proceed in effect as if the recursions are

$$
\begin{aligned}
N_{t-1}^{(0)} &= L_t^{(0)\prime} N_t^{(0)} L_t^{(0)},\\
N_{t-1}^{(1)} &= Z_t' F_t^{(1)} Z_t + L_t^{(0)\prime} N_t^{(1)} L_t^{(0)} + L_t^{(1)\prime} N_t^{(0)} L_t^{(0)},\\
N_{t-1}^{(2)} &= Z_t' F_t^{(2)} Z_t + L_t^{(0)\prime} N_t^{(2)} L_t^{(0)} + L_t^{(0)\prime} N_t^{(1)} L_t^{(1)} + L_t^{(1)\prime} N_t^{(1)\prime} L_t^{(0)} + L_t^{(1)\prime} N_t^{(0)} L_t^{(1)},
\end{aligned} \tag{5.29}
$$

and that we can compute V_t by

$$
V_t = P_{*,t} - P_{*,t} N_{t-1}^{(0)} P_{*,t} - \big(P_{\infty,t} N_{t-1}^{(1)} P_{*,t}\big)' - P_{\infty,t} N_{t-1}^{(1)} P_{*,t} - P_{\infty,t} N_{t-1}^{(2)} P_{\infty,t}. \tag{5.30}
$$

Thus the matrices calculated from (5.26) can be replaced by the ones computed by (5.29) to obtain the correct value for V_t. The new recursions in (5.29) are convenient since the matrix $L_t^{(2)}$ drops out from our calculations for $N_t^{(2)}$.

Indeed the matrix recursions for $N_t^{(1)}$ and $N_t^{(2)}$ in (5.29) are not the same as the recursions for $N_t^{(1)}$ and $N_t^{(2)}$ in (5.26). Also, it can be noticed that matrix $N_t^{(1)}$ in (5.26) is symmetric while $N_t^{(1)}$ in (5.29) is not symmetric. However, the same notation is employed because $N_t^{(1)}$ is only relevant for computing V_t and matrix $P_{\infty,t}N_{t-1}^{(1)}$ is the same when $N_{t-1}^{(1)}$ is computed by (5.26) as when it is computed by (5.29). The same argument applies to matrix $N_t^{(2)}$.

It can now be easily verified that equations (5.30) and the modified recursions (5.29) can be reformulated as

$$N_{t-1}^{\dagger} = \begin{bmatrix} 0 & Z_t'F_t^{(1)}Z_t \\ Z_t'F_t^{(1)}Z_t & Z_t'F_t^{(2)}Z_t \end{bmatrix} + L_t^{\dagger\prime}N_t^{\dagger}L_t^{\dagger}, \qquad V_t = P_{*,t} - P_t^{\dagger}N_{t-1}^{\dagger}P_t^{\dagger\prime}, \tag{5.31}$$

for $t = d, \ldots, 1$, where

$$N_{t-1}^{\dagger} = \begin{bmatrix} N_{t-1}^{(0)} & N_{t-1}^{(1)\prime} \\ N_{t-1}^{(1)} & N_{t-1}^{(2)} \end{bmatrix}, \quad \text{with} \quad N_d^{\dagger} = \begin{bmatrix} N_d & 0 \\ 0 & 0 \end{bmatrix},$$

and the partioned matrices $P_t^{\dagger}$ and $L_t^{\dagger}$ are defined in (5.18). Again, this formulation has the same form as the standard smoothing recursion (4.43) which is a useful property when writing software. The formulations (5.24) and (5.31) are given by Koopman and Durbin (2003).

5.4 Exact initial disturbance smoothing

Calculating the smoothed disturbances does not require as much computing as calculating the smoothed state vector when the initial state vector is diffuse. This is because the smoothed disturbance equations do not involve matrix multiplications which depend on terms in κ or higher order terms. From (4.58) the smoothed estimator is $\hat{\varepsilon}_t = H_t(F_t^{-1}v_t - K_t'r_t)$ where $F_t^{-1} = O(\kappa^{-1})$ for $F_{\infty,t}$ positive definite, $F_t^{-1} = F_{*,t}^{-1} + O(\kappa^{-1})$ for $F_{\infty,t} = 0$, $K_t = K_t^{(0)} + O(\kappa^{-1})$ and $r_t = r_t^{(0)} + O(\kappa^{-1})$ so, as $\kappa \to \infty$, we have

$$\hat{\varepsilon}_t = \begin{cases} -H_tK_t^{(0)\prime}r_t^{(0)} & \text{if } F_{\infty,t} \text{ is nonsingular,} \\ H_t\big(F_{*,t}^{-1}v_t - K_t^{(0)\prime}r_t^{(0)}\big) & \text{if } F_{\infty,t} = 0, \end{cases}$$

for $t = d, \ldots, 1$. Other results for disturbance smoothing are obtained in a similar way and for convenience we collect them together in the form

$$\begin{aligned}
\hat{\varepsilon}_t &= -H_t K_t^{(0)\prime} r_t^{(0)}, \\
\hat{\eta}_t &= Q_t R_t' r_t^{(0)}, \\
\mathrm{Var}(\varepsilon_t | Y_n) &= H_t - H_t K_t^{(0)\prime} N_t^{(0)} K_t^{(0)} H_t, \\
\mathrm{Var}(\eta_t | Y_n) &= Q_t - Q_t R_t' N_t^{(0)} R_t Q_t,
\end{aligned}$$

for the case where $F_{\infty,t} \neq 0$ and

$$\begin{aligned}
\hat{\varepsilon}_t &= H_t \big(F_{*,t}^{-1} v_t - K_t^{(0)\prime} r_t^{(0)} \big), \\
\hat{\eta}_t &= Q_t R_t' r_t^{(0)}, \\
\mathrm{Var}(\varepsilon_t | Y_n) &= H_t - H_t \big(F_{*,t}^{-1} + K_t^{(0)\prime} N_t^{(0)} K_t^{(0)} \big) H_t, \\
\mathrm{Var}(\eta_t | Y_n) &= Q_t - Q_t R_t' N_t^{(0)} R_t Q_t,
\end{aligned}$$

for the case $F_{\infty,t} = 0$ and for $t = d, \ldots, 1$. It is fortunate that disturbance smoothing in the diffuse case does not require as much extra computing as for state smoothing. This is particularly convenient when the score vector is computed repeatedly within the process of parameter estimation as we will discuss in Subsection 7.3.3.

5.5 Exact initial simulation smoothing

5.5.1 Modifications for diffuse initial conditions

When the initial state vector is diffuse it turns out that the simulation smoother of Section 4.9 can be used without the complexities of Section 5.3 required for diffuse state smoothing. Let us begin by looking at diffuse filtering and smoothing from an intuitive point of view. Suppose we were to initialise model (4.12) with an entirely arbitrary value $\alpha_1 = \alpha_1^*$ and then apply formulae in Sections 5.2 to 5.4 to obtain diffuse filtered and smoothed values for the resulting observational vector Y_n. It seems evident intuitively that the filtered and smoothed values that emerge cannot depend on the value of α_1^* that we have chosen.

These ideas suggest the following conjecture for the exact treatment of simulation smoothing in the diffuse case. Set the diffuse elements of α_1 equal to arbitrary values, say zeros, and use the diffuse filters and smoothers developed in Sections 5.2 to 5.4 for the calculation of $\widetilde{w}$ and $\widetilde{\alpha}$ by the methods of Section 4.9, respectively; then these are the exact values required. The validity of this conjecture is proved in Appendix 2 of Durbin and Koopman (2002); details are intricate so they will not be repeated here.

5.5.2 Exact initial simulation smoothing

We first consider how to obtain a simulated sample of α given Y_n using the method of de Jong–Shephard. Taking $\alpha = (\alpha_1, \ldots, \alpha_n, \alpha_{n+1})'$ as before, define $\alpha_{/1}$ as α but without α_1. It follows that

$$p(\alpha|Y_n) = p(\alpha_1|Y_n)p(\alpha_{/1}|Y_n, \alpha_1). \tag{5.32}$$

Obtain a simulated value $\tilde{\alpha}_1$ of α_1 by drawing a sample value from $p(\alpha_1|Y_n) = \mathrm{N}(\hat{\alpha}_1, V_1)$ for which $\hat{\alpha}_1$ and V_1 are computed by the exact initial state smoother as developed in Section 5.3. Next initialise the Kalman filter with $a_1 = \tilde{\alpha}_1$ and $P_1 = 0$, since we now treat $\tilde{\alpha}_1$ as given, and apply the Kalman filter and simulation smoother as decribed in Section 4.9. This procedure for obtaining a sample value of α given Y_n is justified by equation (5.32). To obtain multiple samples, we repeat this procedure. This requires computing a new value of $\tilde{\alpha}_1$, and new values of v_t from the Kalman filter for each new draw. The Kalman filter quantities F_t, K_t and P_{t+1} do not need to be recomputed.

A similar procedure can be followed for simulating disturbance vectors given Y_n: we initialise the Kalman filter with $a_1 = \tilde{\alpha}_1$ and $P_1 = 0$ as above and then use the simulation smoothing recursions of Section 4.9 to generate samples for the disturbances.

5.6 Examples of initial conditions for some models

In this section we give some examples of the exact initial Kalman filter for $t = 1, \ldots, d$ for a range of state space models.

5.6.1 Structural time series models

Structural time series models are usually set up in terms of nonstationary components. Therefore, most of the models in this class have the initial state vector equals δ, that is, $\alpha_1 = \delta$ so that $a_1 = 0$, $P_* = 0$ and $P_\infty = I_m$. We then proceed with the algorithms provided by Sections 5.2, 5.3 and 5.4.

To illustrate the exact initial Kalman filter in detail we consider the local linear trend model (3.2) with system matrices

$$Z_t = (1 \quad 0), \qquad T_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \qquad Q_t = \sigma_\varepsilon^2 \begin{bmatrix} q_\xi & 0 \\ 0 & q_\zeta \end{bmatrix},$$

and with $H_t = \sigma_\varepsilon^2$, $R_t = I_2$, where $q_\xi = \sigma_\xi^2/\sigma_\varepsilon^2$ and $q_\zeta = \sigma_\zeta^2/\sigma_\varepsilon^2$. The exact initial Kalman filter is started with

$$a_1 = 0, \qquad P_{*,1} = 0, \qquad P_{\infty,1} = I_2,$$

and the first update is based on

$$K_1^{(0)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad L_1^{(0)} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \qquad K_1^{(1)} = -\sigma_\varepsilon^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$
$$L_1^{(1)} = \sigma_\varepsilon^2 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

such that

$$a_2 = \begin{pmatrix} y_1 \\ 0 \end{pmatrix}, \qquad P_{*,2} = \sigma_\varepsilon^2 \begin{bmatrix} 1 + q_\xi & 0 \\ 0 & q_\zeta \end{bmatrix}, \qquad P_{\infty,2} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

The second update gives the quantities

$$K_2^{(0)} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \qquad L_2^{(0)} = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix},$$

and

$$K_2^{(1)} = -\sigma_\varepsilon^2 \begin{pmatrix} 3 + q_\xi \\ 2 + q_\xi \end{pmatrix}, \qquad L_2^{(1)} = \sigma_\varepsilon^2 \begin{bmatrix} 3 + q_\xi & 0 \\ 2 + q_\xi & 0 \end{bmatrix},$$

together with the state update results

$$a_3 = \begin{pmatrix} 2y_2 - y_1 \\ y_2 - y_1 \end{pmatrix}, \qquad P_{*,3} = \sigma_\varepsilon^2 \begin{bmatrix} 5 + 2q_\xi + q_\zeta & 3 + q_\xi + q_\zeta \\ 3 + q_\xi + q_\zeta & 2 + q_\xi + 2q_\zeta \end{bmatrix},$$
$$P_{\infty,3} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

It follows that the usual Kalman filter (4.24) can be used for $t = 3, \dots, n$.

5.6.2 Stationary ARMA models

The univariate stationary ARMA model with zero mean of order p and q is given by

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \zeta_t + \theta_1 \zeta_{t-1} + \cdots + \theta_q \zeta_{t-q}, \quad \zeta_t \sim \mathrm{N}(0, \sigma^2).$$

The state space form is

$$y_t = (1, 0, \dots, 0)\alpha_t,$$
$$\alpha_{t+1} = T\alpha_t + R\zeta_{t+1},$$

where the system matrices T and R are given by (3.20) with $r = \max(p, q+1)$. All elements of the state vector are stationary so that the part $a + A\delta$ in (5.2) is zero and $R_0 = I_m$. The unconditional distribution of the initial state vector α_1 is therefore given by

$$\alpha_1 \sim \mathrm{N}(0, \sigma^2 Q_0),$$

where, since $\mathrm{Var}(\alpha_{t+1}) = \mathrm{Var}(T\alpha_t + R\zeta_{t+1})$, $\sigma^2 Q_0 = \sigma^2 T Q_0 T' + \sigma^2 RR'$. This equation needs to be solved for Q_0. It can be shown that a solution can be obtained by solving the linear equation $(I_{m^2} - T \otimes T)\,\mathrm{vec}(Q_0) = \mathrm{vec}(RR')$ for Q_0, where $\mathrm{vec}(Q_0)$ and $\mathrm{vec}(RR')$ are the stacked columns of Q_0 and RR' and where

$$T \otimes T = \begin{bmatrix} t_{11}T & \dots & t_{1m}T \\ t_{21}T & \dots & t_{2m}T \\ \vdots & & \\ t_{m1}T & \dots & t_{mm}T \end{bmatrix},$$

with t_{ij} denoting the (i, j) element of matrix T; see, for example, Magnus and Neudecker (1988) who give a general treatment of problems of this type. The Kalman filter is initialised by $a_1 = 0$ and $P_1 = Q_0$.

As an example, consider the ARMA$(1, 1)$ model

$$y_t = \phi y_{t-1} + \zeta_t + \theta\zeta_{t-1}, \qquad \zeta_t \sim \mathrm{N}(0, \sigma^2).$$

Then

$$T = \begin{bmatrix} \phi & 1 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad R = \begin{pmatrix} 1 \\ \theta \end{pmatrix},$$

so the solution is

$$Q_0 = \begin{bmatrix} (1-\phi^2)^{-1}(1+\theta^2+2\phi\theta) & \theta \\ \theta & \theta^2 \end{bmatrix}.$$

5.6.3 Nonstationary ARIMA models

The univariate nonstationary ARIMA model of order p, d and q can be put in the form

$$y_t^* = \phi_1 y_{t-1}^* + \dots + \phi_p y_{t-p}^* + \zeta_t + \theta_1\zeta_{t-1} + \dots + \theta_q\zeta_{t-q}, \qquad \zeta_t \sim \mathrm{N}(0, \sigma^2).$$

where $y_t^* = \Delta^d y_t$. The state space form of the ARIMA model with $p = 2$, $d = 1$ and $q = 1$ is given in Section 3.4 with the state vector given by

$$\alpha_t = \begin{pmatrix} y_{t-1} \\ y_t^* \\ \phi_2 y_{t-1}^* + \theta_1\zeta_t \end{pmatrix},$$

where $y_t^* = \Delta y_t = y_t - y_{t-1}$. The first element of the initial state vector α_1, that is y_0, is nonstationary while the other elements are stationary. Therefore, the initial vector $\alpha_1 = a + A\delta + R_0\eta_0$ is given by

$$\alpha_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \delta + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \eta_0, \qquad \eta_0 \sim \mathrm{N}(0, Q_0),$$

where Q_0 is the 2×2 unconditional variance matrix for an ARMA model with $p = 2$ and $q = 1$ which we obtain from Subsection 5.6.2. When δ is diffuse, the mean vector and variance matrix are

$$a_1 = 0, \qquad P_1 = \kappa P_\infty + P_*,$$

where

$$P_\infty = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad P_* = \begin{bmatrix} 0 & 0 \\ 0 & Q_0 \end{bmatrix}.$$

Analysis then proceeds using the exact initial Kalman filter and smoother of Sections 5.2, 5.3 and 5.4. The initial state specification for ARIMA models with $d = 1$ but with other values for p and q is obtained in a similar way.

From Section 3.4, the initial state vector for the ARIMA model with $p = 2$, $d = 2$ and $q = 1$ is given by

$$\alpha_1 = \begin{pmatrix} y_0 \\ \Delta y_0 \\ y_1^* \\ \phi_2 y_0^* + \theta_1 \zeta_1 \end{pmatrix}.$$

The first two elements of α_1, that is, y_0 and Δy_0, are nonstationary and we therefore treat them as diffuse. Thus we write

$$\alpha_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \delta + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \eta_0, \qquad \eta_0 \sim \mathrm{N}(0, Q_0),$$

where Q_0 is as for the previous case. It follows that the mean vector and variance matrix of α_1 are

$$a_1 = 0, \qquad P_1 = \kappa P_\infty + P_*,$$

where

$$P_\infty = \begin{bmatrix} I_2 & 0 \\ 0 & 0 \end{bmatrix}, \qquad P_* = \begin{bmatrix} 0 & 0 \\ 0 & Q_0 \end{bmatrix}.$$

We then proceed with the methods of Sections 5.2, 5.3 and 5.4. The initial conditions for non-seasonal ARIMA models with other values for p, d and q and seasonal models are derived in similar ways.

5.6.4 Regression model with ARMA errors

The regression model with k explanatory variables and ARMA(p, q) errors (3.30) can be written in state space form as in Subsection 3.6.2. The initial state vector is

$$\alpha_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{bmatrix} I_k \\ 0 \end{bmatrix} \delta + \begin{bmatrix} 0 \\ I_r \end{bmatrix} \eta_0, \qquad \eta_0 \sim \mathrm{N}(0, Q_0),$$

where Q_0 is obtained as in Subsection 5.6.2 and $r = \max(p, q+1)$. When δ is treated as diffuse we have $\alpha_1 \sim \mathrm{N}(a_1, P_1)$ where $a_1 = 0$ and $P_1 = \kappa P_\infty + P_*$ with

$$P_\infty = \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix}, \qquad P_* = \begin{bmatrix} 0 & 0 \\ 0 & Q_0 \end{bmatrix}.$$

We then proceed as described in the last section.

To illustrate the use of the exact initial Kalman filter we consider the simple case of an AR(1) model with a constant, that is

$$\begin{aligned} y_t &= \mu + \xi_t, \\ \xi_t &= \phi \xi_{t-1} + \zeta_t, \qquad \zeta_t \sim \mathrm{N}(0, \sigma^2). \end{aligned}$$

In state space form we have

$$\alpha_t = \begin{pmatrix} \mu \\ \xi_t \end{pmatrix}$$

and the system matrices are given by

$$Z_t = (1 \quad 1), \qquad T_t = \begin{bmatrix} 1 & 0 \\ 0 & \phi \end{bmatrix}, \qquad R_t = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

with $H_t = 0$ and $Q_t = \sigma^2$. The exact initial Kalman filter is started with

$$a_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad P_{*,1} = c \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \qquad P_{\infty,1} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

where $c = \sigma^2/(1 - \phi^2)$. The first update is based on

$$K_1^{(0)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad L_1^{(0)} = \begin{bmatrix} 0 & -1 \\ 0 & \phi \end{bmatrix}, \qquad K_1^{(1)} = c \begin{pmatrix} -1 \\ \phi \end{pmatrix},$$
$$L_1^{(1)} = c \begin{bmatrix} 1 & 1 \\ -\phi & -\phi \end{bmatrix},$$

such that

$$a_2 = \begin{pmatrix} y_1 \\ 0 \end{pmatrix}, \qquad P_{*,2} = \frac{\sigma^2}{1 - \phi^2} \begin{bmatrix} 1 & -\phi \\ -\phi & 1 \end{bmatrix}, \qquad P_{\infty,2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

It follows that the usual Kalman filter (4.24) can be used for $t = 2, \ldots, n$.

5.6.5 Spline smoothing

The initial state vector for the spline model (3.44) is simply $\alpha_1 = \delta$, implying that $a_1 = 0$, $P_* = 0$ and $P_\infty = I_2$.

5.7 Augmented Kalman filter and smoother

5.7.1 Introduction

An alternative approach for dealing with the initialisation problem is owed to Rosenberg (1973), de Jong (1988b) and de Jong (1991). As in (5.2), the initial state vector is defined as

$$\alpha_1 = a + A\delta + R_0\eta_0, \qquad \eta_0 \sim \mathrm{N}(0, Q_0). \tag{5.33}$$

Rosenberg (1973) treats δ as a fixed unknown vector and he employs maximum likelihood to estimate δ while de Jong (1991) treats δ as a diffuse vector. Since the treatments of Rosenberg and de Jong are both based on the idea of augmenting the observed vector, we will refer to their procedures collectively as the *augmentation approach.* The approach of Rosenberg offers relief to analysts who feel uncomfortable about using diffuse initialising densities on the ground that infinite variances have no counterpart in real data. In fact, as we shall show, the two approaches give effectively the same answer so these analysts could regard the diffuse assumption as a device for achieving initialisation based on maximum likelihood estimates of the unknown initial state elements. The results are developed from a classical point of view for the case where the observations are normally distributed. Corresponding results for MVLUE and Bayesian analyses can be obtained by applying Lemmas 2, 3 and 4 of Section 4.2.

5.7.2 Augmented Kalman filter

In this subsection we establish the groundwork for both the Rosenberg and the de Jong techniques. For given δ, apply the Kalman filter, (4.24) with $a_1 = \mathrm{E}(\alpha_1) = a + A\delta$, $P_1 = \mathrm{Var}(\alpha_1) = P_* = R_0Q_0R_0'$ and denote the resulting value of a_t from the filter output by $a_{\delta,t}$. Since $a_{\delta,t}$ is a linear function of the observations and $a_1 = a + A\delta$ we can write

$$a_{\delta,t} = a_{a,t} + A_{A,t}\delta, \tag{5.34}$$

where $a_{a,t}$ is the value of a_t in the filter output obtained by taking $a_1 = a$, $P_1 = P_*$ and where the jth column of $A_{A,t}$ is the value of a_t in the filter output obtained from an observational vector of zeros with $a_1 = A_j$, $P_1 = P_*$, where A_j is the jth column of A. Denote the value of v_t in the filter output obtained by taking $a_1 = a + A\delta$, $P_1 = P_*$ by $v_{\delta,t}$. Analogously to (5.34) we can write

$$v_{\delta,t} = v_{a,t} + V_{A,t}\delta, \tag{5.35}$$

where $v_{a,t}$ and $V_{A,t}$ are given by the same Kalman filters that produced $a_{a,t}$ and $A_{A,t}$.

The matrices $(a_{a,t}, A_{A,t})$ and $(v_{a,t}, V_{A,t})$ can be computed in one pass through a Kalman filter which inputs the observation vector y_t augmented by zero values. This is possible because for each Kalman filter producing a particular column of $(a_{a,t}, A_{A,t})$, the same variance initialisation $P_1 = P_*$ applies, so the variance output, which we denote by $F_{\delta,t}$, $K_{\delta,t}$ and $P_{\delta,t+1}$, is the same for each Kalman filter. Replacing the Kalman filter equations for the vectors v_t and a_t by the corresponding equations for the matrices $(a_{a,t}, A_{A,t})$ and $(v_{a,t}, V_{A,t})$ leads to the equations

$$\begin{aligned} (v_{a,t}, V_{A,t}) &= (y_t, 0) - Z_t(a_{a,t}, A_{A,t}), \\ (a_{a,t+1}, A_{A,t+1}) &= T_t(a_{a,t}, A_{A,t}) + K_{\delta,t}(v_{a,t}, V_{A,t}), \end{aligned} \tag{5.36}$$

where $(a_{a,1}, A_{A,1}) = (a, A)$; the recursions corresponding to F_t, K_t and P_{t+1} remain as for the standard Kalman filter, that is,

$$\begin{aligned} F_{\delta,t} &= Z_t P_{\delta,t} Z_t' + H_t, \\ K_{\delta,t} &= T_t P_{\delta,t} Z_t' F_{\delta,t}^{-1}, \qquad L_{\delta,t} = T_t - K_{\delta,t} Z_t, \\ P_{\delta,t+1} &= T_t P_{\delta,t} L_{\delta,t}' + R_t Q_t R_t', \end{aligned} \tag{5.37}$$

for $t = 1, \ldots, n$ with $P_{\delta,1} = P_*$. We have included the suffix δ in these expressions not because they depend mathematically on δ but because they have been calculated on the assumption that δ is fixed. The modified Kalman filter (5.36) and (5.37) will be referred to as the *augmented Kalman filter* in this book.

5.7.3 Filtering based on the augmented Kalman filter

With these preliminaries, let us first consider the diffuse case (5.2) with $\delta \sim \mathrm{N}(0, \kappa I_q)$ where $\kappa \to \infty$; we will consider later the case where δ is fixed and is estimated by maximum likelihood. From (5.34) we obtain for given κ,

$$a_{t+1} = \mathrm{E}(\alpha_{t+1}|Y_t) = a_{a,t+1} + A_{A,t+1}\bar{\delta}_t, \tag{5.38}$$

where $\bar{\delta}_t = \mathrm{E}(\delta|Y_t)$. Now

$$\begin{aligned} \log p(\delta|Y_t) &= \log p(\delta) + \log p(Y_t|\delta) - \log p(Y_t) \\ &= \log p(\delta) + \sum_{j=1}^{t} \log p(v_{\delta,j}) - \log p(Y_t) \\ &= -\frac{1}{2\kappa}\delta'\delta - b_t'\delta - \frac{1}{2}\delta' S_{A,t}\delta + \text{terms independent of } \delta, \end{aligned} \tag{5.39}$$

where

$$b_t = \sum_{j=1}^{t} V_{A,j}' F_{\delta,j}^{-1} v_{a,j}, \qquad S_{A,t} = \sum_{j=1}^{t} V_{A,j}' F_{\delta,j}^{-1} V_{A,j}. \tag{5.40}$$

Since densities are normal, the mean of $p(\delta|Y_t)$ is equal to the mode, and this is the value of δ which maximises $\log p(\delta|Y_t)$, so on differentiating (5.39) with respect to δ and equating to zero, we have

$$\bar{\delta}_t = -\left(S_{A,t} + \frac{1}{\kappa}I_q\right)^{-1} b_t. \tag{5.41}$$

Also,

$$\begin{aligned} P_{t+1} &= \mathrm{E}[(a_{t+1} - \alpha_{t+1})(a_{t+1} - \alpha_{t+1})'] \\ &= \mathrm{E}[\{a_{\delta,t+1} - \alpha_{t+1} - A_{A,t+1}(\delta - \bar{\delta}_t)\}\{a_{\delta,t+1} - \alpha_{t+1} - A_{A,t+1}(\delta - \bar{\delta}_t)\}'] \\ &= P_{\delta,t+1} + A_{A,t+1}\mathrm{Var}(\delta|Y_t)A'_{A,t+1} \\ &= P_{\delta,t+1} + A_{A,t+1}\left(S_{A,t} + \frac{1}{\kappa}I_q\right)^{-1} A'_{A,t+1}, \end{aligned} \tag{5.42}$$

since $\mathrm{Var}(\delta|Y_t) = (S_{A,t} + \frac{1}{\kappa}I_q)^{-1}$.

Letting $\kappa \to \infty$ we have

$$\bar{\delta}_t = -S_{A,t}^{-1}b_t, \tag{5.43}$$

$$\mathrm{Var}(\delta|Y_t) = S_{A,t}^{-1}, \tag{5.44}$$

when $S_{A,t}$ is nonsingular. The calculations of b_t and $S_{A,t}$ are easily incorporated into the augmented Kalman filter (5.36) and (5.37). It follows that

$$a_{t+1} = a_{a,t+1} - A_{A,t+1}S_{A,t}^{-1}b_t, \tag{5.45}$$

$$P_{t+1} = P_{\delta,t+1} + A_{A,t+1}S_{A,t}^{-1}A'_{A,t+1}, \tag{5.46}$$

as $\kappa \to \infty$. For $t < d$, $S_{A,t}$ is singular so values of a_{t+1} and P_{t+1} given by (5.45) and (5.46) do not exist. However, when $t = d$, a_{d+1} and P_{d+1} exist and consequently when $t > d$ the values a_{t+1} and P_{t+1} can be calculated by the standard Kalman filter for $t = d+1, \ldots, n$. Thus we do not need to use the augmented Kalman filter (5.36) for $t = d+1, \ldots, n$. These results are due to de Jong (1991) but our derivation here is more transparent.

We now consider a variant of the maximum likelihood method for initialising the filter due to Rosenberg (1973). In this technique, δ is regarded as fixed and unknown and we employ maximum likelihood given Y_t to obtain estimate $\hat{\delta}_t$. The loglikelihood of Y_t given δ is

$$\log p(Y_t|\delta) = \sum_{j=1}^{t} \log p(v_{\delta,j}) = -b_t'\delta - \frac{1}{2}\delta' S_{A,t}\delta + \text{terms independent of } \delta,$$

which is the same as (5.39) apart from the term $-\delta'\delta/(2\kappa)$. Differentiating with respect to δ, equating to zero and taking the second derivative gives

$$\hat{\delta}_t = -S_{A,t}^{-1} b_t, \qquad \mathrm{Var}(\hat{\delta}_t) = S_{A,t}^{-1},$$

when $S_{A,t}$ is nonsingular, that is for $t = d, \ldots, n$. These values are the same as $\bar{\delta}_t$ and $\mathrm{Var}(\delta_t|Y_t)$ when $\kappa \to \infty$. In practice we choose t to be the smallest value for which $\hat{\delta}_t$ exists, which is d. It follows that the values of a_{t+1} and P_{t+1} for $t \geq d$ given by this approach are the same as those obtained in the diffuse case. Thus the solution of the initialisation problem given in Section 5.2 also applies to the case where δ is fixed and unknown. From a computational point of view the calculations of Section 5.2 are more efficient than those for the augmented device described in this section when the model is reasonably large. A comparison of the computational efficiency is given in Subsection 5.7.5. Rosenberg (1973) used a procedure which differed slightly from this. Although he employed essentially the same augmentation technique, in the notation above he estimated δ by the value $\hat{\delta}_n$ based on all the data.

5.7.4 Illustration: the local linear trend model

To illustrate the augmented Kalman filter we consider the same local linear trend model as in Subsection 5.6.1. The system matrices of the local linear trend model (3.2) are given by

$$Z = (1 \quad 0), \qquad T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \qquad Q = \sigma_\varepsilon^2 \begin{bmatrix} q_\xi & 0 \\ 0 & q_\zeta \end{bmatrix},$$

with $H = \sigma_\varepsilon^2$ and $R = I_2$ and where $q_\xi = \sigma_\xi^2/\sigma_\varepsilon^2$ and $q_\zeta = \sigma_\zeta^2/\sigma_\varepsilon^2$. The augmented Kalman filter is started with

$$(a_{a,1}, A_{A,1}) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad P_{\delta,1} = \sigma_\varepsilon^2 \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

and the first update is based on

$$(v_{a,1}, V_{A,1}) = (y_1 \quad -1 \quad 0), \qquad F_{\delta,1} = \sigma_\varepsilon^2, \qquad K_{\delta,1} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$$L_{\delta,1} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix},$$

so

$$b_1 = -\frac{1}{\sigma_\varepsilon^2} \begin{pmatrix} y_1 \\ 0 \end{pmatrix}, \qquad S_{A,1} = \frac{1}{\sigma_\varepsilon^2} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

and

$$(a_{a,2}, A_{A,2}) = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \qquad P_{\delta,2} = \sigma_\varepsilon^2 \begin{bmatrix} q_\xi & 0 \\ 0 & q_\zeta \end{bmatrix}.$$

The second update gives the quantities

$$(v_{a,2}, V_{A,2}) = (y_2 \quad -1 \quad -1), \qquad F_{\delta,2} = \sigma_\varepsilon^2(1+q_\xi),$$
$$K_{\delta,2} = \frac{q_\xi}{1+q_\xi}\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad L_{\delta,2} = \begin{bmatrix} \frac{1}{1+q_\xi} & 1 \\ 0 & 1 \end{bmatrix},$$

with

$$b_2 = \frac{-1}{1+q_\xi}\begin{pmatrix} (1+q_\xi)y_1 + y_2 \\ y_2 \end{pmatrix}, \qquad S_{A,2} = \frac{1}{\sigma_\varepsilon^2(1+q_\xi)}\begin{bmatrix} 2+q_\xi & 1 \\ 1 & 1 \end{bmatrix},$$

and the state update results

$$(a_{a,3}, A_{A,3}) = \frac{1}{1+q_\xi}\begin{bmatrix} q_\xi y_2 & 1 & 2+q_\xi \\ 0 & 0 & 1+q_\xi \end{bmatrix}, \quad P_{\delta,3} = \sigma_\varepsilon^2 \begin{bmatrix} q_\xi + \frac{q_\xi}{1+q_\xi} + q_\zeta & q_\zeta \\ q_\zeta & 2q_\zeta \end{bmatrix}.$$

The augmented part can be collapsed since $S_{A,2}$ is nonsingular, giving

$$S_{A,2}^{-1} = \sigma_\varepsilon^2 \begin{bmatrix} 1 & -1 \\ -1 & 2+q_\xi \end{bmatrix}, \qquad \bar{\delta}_2 = -S_{A,2}^{-1} b_2 = \begin{pmatrix} y_1 \\ y_2 - y_1 \end{pmatrix}.$$

It follows that

$$a_3 = a_{a,3} + A_{A,3}\bar{\delta}_2 = \begin{pmatrix} 2y_2 - y_1 \\ y_2 - y_1 \end{pmatrix},$$
$$P_3 = P_{\delta,3} + A_{A,3} S_{A,2}^{-1} A'_{A,3} = \sigma_\varepsilon^2 \begin{bmatrix} 5 + 2q_\xi + q_\zeta & 3 + q_\xi + q_\zeta \\ 3 + q_\xi + q_\zeta & 2 + q_\xi + 2q_\zeta \end{bmatrix}.$$

and the usual Kalman filter (4.24) can be used for $t = 3, \ldots, n$. These results are exactly the same as those obtained in Section 5.6.1, though the computations take longer as we will now show.

5.7.5 Comparisons of computational efficiency

The adjusted Kalman filters of Section 5.2 and Subsection 5.7.2 both require more computations than the Kalman filter (4.24) with known initial conditions. Of course the adjustments are only required for a limited number of updates. The additional computations for the exact initial Kalman filter are due to updating the matrix $P_{\infty,t+1}$ and computing the matrices $K_t^{(1)}$ and $L_t^{(1)}$ when $F_{\infty,t} \neq 0$, for $t = 1, \ldots, d$. For many practical state space models the system matrices Z_t and T_t are sparse selection matrices containing many zeros and ones; this is the case for the models discussed in Chapter 3. Therefore, calculations involving Z_t and T_t are particularly cheap for most models. Table 5.1 compares the number of additional multiplications (compared to the Kalman filter with known initial

Table 5.1 Number of additional multiplications for filtering.

Model	Exact initial	Augmenting	Difference (%)
Local level	3	7	57
Local linear trend	18	46	61
Basic seasonal (s = 4)	225	600	63
Basic seasonal (s = 12)	3549	9464	63

conditions) required for filtering using the devices of Section 5.2 and Subsection 5.7.2 applied to several structural time series models which are discussed in Section 3.2. The results in Table 5.1 show that the additional number of computations for the exact initial Kalman filter of Section 5.2 is less than half the extra computations required for the augmentation device of Subsection 5.7.2. Such computational efficiency gains are important when the Kalman filter is used many times as is the case for parameter estimation; a detailed discussion of estimation is given in Chapter 7. It will also be argued in Subsection 7.3.5 that many computations for the exact initial Kalman filter only need to be done once for a specific model since the computed values remain the same when the parameters of the model change. This argument does not apply to the augmentation device and this is another important reason why our approach in Section 5.2 is more efficient than the augmentation approach.

5.7.6 Smoothing based on the augmented Kalman filter

The smoothing algorithms can also be developed using the augmented approach. The smoothing recursion for r_{t-1} in (4.69) needs to be augmented in the same way as is done for v_t and a_t of the Kalman filter. When the augmented Kalman filter is applied for $t = 1, \dots, n$, the modifications for smoothing are straightforward after computing $\hat{\delta}_n$ and $\mathrm{Var}(\hat{\delta}_n)$ and then applying similar expressions to those of (5.45) and (5.46). The collapse of the augmented Kalman filter to the standard Kalman filter is computationally efficient for filtering but, as a result, the estimates $\hat{\delta}_n$ and $\mathrm{Var}(\hat{\delta}_n)$ are not available for calculating the smoothed estimates of the state vector. It is not therefore straightforward to do smoothing when the collapsing device is used in the augmentation approach. A solution for this problem has been given by Chu-Chun-Lin and de Jong (1993).