4 Filtering, smoothing and forecasting

4.1 Introduction

In this chapter and the following three chapters we provide a general treatment from both classical and Bayesian perspectives of the linear Gaussian state space model (3.1). The observations y_t will be treated as multivariate. For much of the theory, the development is a straightforward extension to the general case of the treatment of the simple local level model in Chapter 2. We also consider linear unbiased estimates in the non-normal case.

In Section 4.2 we present some elementary results in multivariate regression theory which provide the foundation for our treatment of Kalman filtering and smoothing later in the chapter. We begin by considering a pair of jointly distributed random vectors x and y. Assuming that their joint distribution is normal, we show in Lemma 1 that the conditional distribution of x given y is normal and we derive its mean vector and variance matrix. We shall show in Section 4.3 that these results lead directly to the Kalman filter. For workers who do not wish to assume normality we derive in Lemma 2 the minimum variance linear unbiased estimate of x given y. For those who prefer a Bayesian approach we derive in Lemma 3, under the assumption of normality, the posterior density of x given an observed value of y. Finally in Lemma 4, while retaining the Bayesian approach, we drop the assumption of normality and derive a quasi-posterior density of x given y, with a mean vector which is linear in y and which has minimum variance matrix.

All four lemmas can be regarded as representing in appropriate senses the regression of x on y. For this reason in all cases the mean vectors and variance matrices are the same. We shall use these lemmas to derive the Kalman filter and smoother in Sections 4.3 and 4.4. Because the mean vectors and variance matrices are the same, we need only use one of the four lemmas to derive the results that we need; the results so obtained then remain valid under the conditions assumed under the other three lemmas.

Denote the set of observations $y_1, \ldots, y_t$ by Y_t. In Section 4.3 we will derive the Kalman filter, which is a recursion for calculating $a_{t|t} = \mathrm{E}(\alpha_t|Y_t)$, $a_{t+1} = \mathrm{E}(\alpha_{t+1}|Y_t)$, $P_{t|t} = \mathrm{Var}(\alpha_t|Y_t)$ and $P_{t+1} = \mathrm{Var}(\alpha_{t+1}|Y_t)$ given a_t and P_t. The derivation requires only elementary properties of multivariate regression theory derived in Lemmas 1 to 4. We also investigate some properties of state estimation errors and one-step ahead forecast errors. In Section 4.4 we use the output of the Kalman filter and the properties of forecast errors to obtain recursions for smoothing the series, that is, calculating the conditional mean and variance matrix of α_t, for $t = 1, \ldots, n, n+1$, given all the observations $y_1, \ldots, y_n$. Estimates of the disturbance vectors ε_t and η_t and their error variance matrices given all the data are derived in Section 4.5. Covariance matrices of smoothed estimators are considered in Section 4.7. The weights associated with filtered and smoothed estimates of functions of the state and disturbance vectors are discussed in Section 4.8. Section 4.9 describes how to generate random samples for purposes of simulation from the smoothed densities of the state and disturbance vectors given the observations. The problem of missing observations is considered in Section 4.10 where we show that with the state space approach the problem is easily dealt with by means of simple modifications of the Kalman filter and the smoothing recursions. Section 4.11 shows that forecasts of observations and state can be obtained simply by treating future observations as missing values; these results are of special significance in view of the importance of forecasting in much practical time series work. A comment on varying dimensions of the observation vector is given in Section 4.12. Finally, in Section 4.13 we consider a general matrix formulation of the state space model.

4.2 Basic results in multivariate regression theory

In this section we present some basic results in elementary multivariate regression theory that we shall use for the development of the theory for the linear Gaussian state space model (3.1) and its non-Gaussian version with $\varepsilon_t \sim (0, H_t)$ and $\eta_t \sim (0, Q_t)$. We shall present the results in a general form before embarking on the state space theory because we shall need to apply them in a variety of different contexts and it is preferable to prove them once only in general rather than to produce a series of similar *ad hoc* proofs tailored to specific situations. A further point is that this form of presentation exposes the intrinsically simple nature of the mathematical theory underlying the state space approach to time series analysis. Readers who are prepared to take for granted the results in Lemmas 1 to 4 below can skip the proofs and go straight to Section 4.3.

Suppose that x and y are jointly normally distributed random vectors with

$$\mathrm{E}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \qquad \mathrm{Var}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma'_{xy} & \Sigma_{yy} \end{bmatrix}, \tag{4.1}$$

where Σ_{yy} is assumed to be a nonsingular matrix.

Lemma 1 *The conditional distribution of x given y is normal with mean vector*

$$\mathrm{E}(x|y) = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y), \tag{4.2}$$

and variance matrix

$$\mathrm{Var}(x|y) = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma'_{xy}. \tag{4.3}$$

Proof. Let $z = x - \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y)$. Since the transformation from (x, y) to (y, z) is linear and (x, y) is normally distributed, the joint distribution of y and z is normal. We have

$$\begin{aligned}
\mathrm{E}(z) &= \mu_x \\
\mathrm{Var}(z) &= \mathrm{E}\left[(z - \mu_x)(z - \mu_x)'\right] \\
&= \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}', \qquad (4.4) \\
\mathrm{Cov}(y, z) &= \mathrm{E}\left[y(z - \mu_x)'\right] \\
&= \mathrm{E}\left[y(x - \mu_x)' - y(y - \mu_y)'\Sigma_{yy}^{-1}\Sigma_{xy}'\right] \\
&= 0. \qquad (4.5)
\end{aligned}$$

Using the result that if two vectors are normal and uncorrelated they are independent, we infer from (4.5) that z is distributed independently of y. Since the distribution of z does not depend on y its conditional distribution given y is the same as its unconditional distribution, that is, it is normal with mean vector μ_x and variance matrix (4.4) which is the same as (4.3). Since $z = x - \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y)$, it follows that the conditional distribution of x given y is normal with mean vector (4.2) and variance matrix (4.3). □

Formulae (4.2) and (4.3) are well known in regression theory. An early proof in a state space context is given in Åström (1970, Chapter 7, Theorem 3.2). The proof given here is based on the treatment given by Rao (1973, §8a.2(v)). A partially similar proof is given by Anderson (2003, Theorem 2.5.1). A quite different proof in a state space context is given by Anderson and Moore (1979, Example 3.2) which is repeated by Harvey (1989, Appendix to Chapter 3); some details of this proof are given in Exercise 4.14.1.

We can regard Lemma 1 as representing the regression of x on y in a multivariate normal distribution. It should be noted that Lemma 1 remains valid when Σ_{yy} is singular if the symbol Σ_{yy}^{-1} is interpreted as a generalised inverse; see the treatment in Rao (1973). Åström (1970) pointed out that if the distribution of (x, y) is singular we can always derive a nonsingular distribution by making a projection on the hyperplanes where the mass is concentrated. The fact that the conditional variance $\mathrm{Var}(x|y)$ given by (4.3) does not depend on y is a property special to the multivariate normal distribution and does not generally hold for other distributions.

We now consider the estimation of x when x is unknown and y is known, as for example when y is an observed vector. Under the assumptions of Lemma 1 we take as our estimate of x the conditional expectation $\widehat{x} = \mathrm{E}(x|y)$, that is,

$$\widehat{x} = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y). \qquad (4.6)$$

This has estimation error $\widehat{x} - x$ so $\widehat{x}$ is conditionally unbiased in the sense that $\mathrm{E}(\widehat{x} - x|y) = \widehat{x} - \mathrm{E}(x|y) = 0$. It is also obviously unconditionally unbiased in the sense that $\mathrm{E}(\widehat{x} - x) = 0$. The unconditional error variance matrix of $\widehat{x}$ is

$$\mathrm{Var}(\widehat{x}-x) = \mathrm{Var}\left[\Sigma_{xy}\Sigma_{yy}^{-1}(y-\mu_y)-(x-\mu_x)\right] \;=\; \Sigma_{xx}-\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}'. \quad (4.7)$$

Expressions (4.6) and (4.7) are, of course, the same as (4.2) and (4.3) respectively.

We now consider the estimation of x given y when the assumption that (x, y) is normally distributed is dropped. We assume that the other assumptions of Lemma 1 are retained. Let us restrict our attention to estimates $\bar{x}$ that are linear in the elements of y, that is, we shall take

$$\bar{x} = \beta + \gamma y,$$

where β is a fixed vector and γ is a fixed matrix of appropriate dimensions. The estimation error is $\bar{x}-x$. If $\mathrm{E}(\bar{x}-x)=0$, we say that $\bar{x}$ is a *linear unbiased estimate* (LUE) of x given y. If there is a particular value x^* of $\bar{x}$ such that

$$\mathrm{Var}(\bar{x}-x)-\mathrm{Var}(x^*-x),$$

is non-negative definite for all LUEs $\bar{x}$ we say that x^* is a *minimum variance linear unbiased estimate* (MVLUE) of x given y. Note that the mean vectors and variance matrices here are unconditional and not conditional given y as were considered in Lemma 1. An MVLUE for the non-normal case is given by the following lemma.

Lemma 2 *Whether (x,y) is normally distributed or not, the estimate $\widehat{x}$ defined by* (4.6) *is a MVLUE of x given y and its error variance matrix is given by* (4.7).

Proof. Since $\bar{x}$ is an LUE, we have

$$\begin{aligned}\mathrm{E}(\bar{x}-x) &= \mathrm{E}(\beta+\gamma y-x)\\ &= \beta+\gamma\mu_y-\mu_x \;=\; 0.\end{aligned}$$

It follows that $\beta = \mu_x - \gamma\mu_y$ and therefore

$$\bar{x} = \mu_x + \gamma(y-\mu_y). \quad (4.8)$$

Thus

$$\begin{aligned}\mathrm{Var}(\bar{x}-x) &= \mathrm{Var}\left[\mu_x+\gamma(y-\mu_y)-x\right]\\ &= \mathrm{Var}\left[\gamma(y-\mu_y)-(x-\mu_x)\right]\\ &= \gamma\Sigma_{yy}\gamma'-\gamma\Sigma_{xy}'-\Sigma_{xy}\gamma'+\Sigma_{xx}\\ &= \mathrm{Var}\left[(\gamma-\Sigma_{xy}\Sigma_{yy}^{-1})y\right]+\Sigma_{xx}-\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}'. \quad (4.9)\end{aligned}$$

Let $\widehat{x}$ be the value of $\bar{x}$ obtained by putting $\gamma = \Sigma_{xy}\Sigma_{yy}^{-1}$ in (4.8). Then $\widehat{x} = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y-\mu_y)$ and from (4.9), it follows that

$$\mathrm{Var}(\widehat{x}-x) = \Sigma_{xx}-\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}'.$$

We can therefore rewrite (4.9) in the form

$$\mathrm{Var}(\bar{x} - x) = \mathrm{Var}\left[(\gamma - \Sigma_{xy}\Sigma_{yy}^{-1})y\right] + \mathrm{Var}(\widehat{x} - x), \tag{4.10}$$

which holds for all LUEs $\bar{x}$. Since $\mathrm{Var}\left[(\gamma - \Sigma_{xy}\Sigma_{yy}^{-1})y\right]$ is non-negative definite the lemma is proved. □

The MVLUE property of the vector estimate $\widehat{x}$ implies that arbitrary linear functions of elements of $\widehat{x}$ are minimum variance linear unbiased estimates of the corresponding linear functions of the elements of x. Lemma 2 can be regarded as an analogue for multivariate distributions of the Gauss–Markov theorem for least squares regression of a dependent variable on fixed regressors. For a treatment of the Gauss–Markov theorem, see, for example, Davidson and MacKinnon (1993, Chapter 3). Lemma 2 is proved in the special context of Kalman filtering by Duncan and Horn (1972) and by Anderson and Moore (1979, §3.2). However, their treatments lack the brevity and generality of Lemma 2 and its proof.

Lemma 2 is highly significant for workers who prefer not to assume normality as the basis for the analysis of time series on the grounds that many real time series have distributions that appear to be far from normal; however, the MVLUE criterion is regarded as acceptable as a basis for analysis by many of these workers. We will show later in the book that many important results in state space analysis such as Kalman filtering and smoothing, missing observation analysis and forecasting can be obtained by using Lemma 1; Lemma 2 shows that these results also satisfy the MVLUE criterion. A variant of Lemma 2 is to formulate it in terms of minimum mean square error matrix rather than minimum variance unbiasedness; this variant is dealt with in Exercise 4.14.4.

Other workers prefer to treat inference problems in state space time series analysis from a Bayesian point of view instead of from the classical standpoint for which Lemmas 1 and 2 are appropriate. We therefore consider basic results in multivariate regression theory that will lead us to a Bayesian treatment of the linear Gaussian state space model.

Suppose that x is a parameter vector with *prior density* $p(x)$ and that y is an observational vector with density $p(y)$ and conditional density $p(y|x)$. Suppose further that the joint density of x and y is the multivariate normal density $p(x, y)$. Then the *posterior density* of x given y is

$$p(x|y) = \frac{p(x, y)}{p(y)} = \frac{p(x)p(y|x)}{p(y)}. \tag{4.11}$$

We shall use the same notation as in (4.1) for the first and second moments of x and y. The equation (4.11) is a form of Bayes Theorem.

Lemma 3 *The posterior density of x given y is normal with posterior mean vector* (4.2) *and posterior variance matrix* (4.3).

Proof. Using (4.11), the proof follows immediately from Lemma 1. □

A general Bayesian treatment of state space time series analysis is given by West and Harrison (1997, §17.2.2) in which an explicit proof of our Lemma 3 is given. Their results emerge in a different form from our (4.2) and (4.3) which, as they point out, can be converted to ours by an algebraical identity; see Exercise 4.14.5 for details.

We now drop the assumption of normality and derive a Bayesian-type analogue of Lemma 2. Let us introduce a broader concept of a posterior density than the traditional one by using the term 'posterior density' to refer to *any* density of x given y and not solely to the conditional density of x given y. Let us consider posterior densities whose mean $\bar{x}$ given y is linear in the elements of y, that is, we take $\bar{x} = \beta + \gamma y$ where β is a fixed vector and γ is a fixed matrix; we say that $\bar{x}$ is a *linear posterior mean.* If there is a particular value x^* of $\bar{x}$ such that

$$\mathrm{Var}(\bar{x} - x) - \mathrm{Var}(x^* - x)$$

is non-negative definite for all linear posterior means $\bar{x}$, we say that x^* is a *minimum variance linear posterior mean estimate* (MVLPME) of x given y.

Lemma 4 *The linear posterior mean $\widehat{x}$ defined by* (4.6) *is a MVLPME and its error variance matrix is given by* (4.7).

Proof. Taking expectations with respect to density $p(y)$ we have

$$\mathrm{E}(\bar{x}) = \mu_x = \mathrm{E}(\beta + \gamma y) = \beta + \gamma \mu_y,$$

from which it follows that $\beta = \mu_x - \gamma \mu_y$ and hence that (4.8) holds. Let $\widehat{x}$ be the value of $\bar{x}$ obtained by putting $\gamma = \Sigma_{xy}\Sigma_{yy}^{-1}$ in (4.8). It follows as in the proof of Lemma 2 that (4.10) applies so the lemma is proved. □

The four lemmas in this section have an important common property. Although each lemma starts from a different criterion, they all finish up with distributions which have the same mean vector (4.2) and the same variance matrix (4.3). The significance of this result is that formulae for the Kalman filter, its associated smoother and related results throughout Part I of the book are exactly the same whether an individual worker wishes to start from a criterion of classical inference, minimum variance linear unbiased estimation or Bayesian inference.

4.3 Filtering

4.3.1 Derivation of the Kalman filter

For convenience we restate the linear Gaussian state space model (3.1) here as

$$\begin{aligned} y_t &= Z_t\alpha_t + \varepsilon_t, & \varepsilon_t &\sim \mathrm{N}(0, H_t), & & \\ \alpha_{t+1} &= T_t\alpha_t + R_t\eta_t, & \eta_t &\sim \mathrm{N}(0, Q_t), & t &= 1, \dots, n, \\ & & \alpha_1 &\sim \mathrm{N}(a_1, P_1), & & \end{aligned} \tag{4.12}$$

where details are given below (3.1). At various points we shall drop the normality assumptions in (4.12). Let Y_{t-1} denote the set of past observations $y_1, \dots, y_{t-1}$ for $t = 2, 3, \dots$ while Y_0 indicates that there is no prior observation before $t = 1$. In our treatments below, we will define Y_t by the vector $(y'_1, \dots, y'_t)'$. Starting at $t = 1$ in (4.12) and building up the distributions of α_t and y_t recursively, it is easy to show that $p(y_t|\alpha_1, \dots, \alpha_t, Y_{t-1}) = p(y_t|\alpha_t)$ and $p(\alpha_{t+1}|\alpha_1, \dots, \alpha_t, Y_t) = p(\alpha_{t+1}|\alpha_t)$. In Table 4.1 we give the dimensions of the vectors and matrices of the state space model.

In this section we derive the Kalman filter for model (4.12) for the case where the initial state α_1 is $\mathrm{N}(a_1, P_1)$ where a_1 and P_1 are known. We shall base the derivation on classical inference using Lemma 1. It follows from Lemmas 2 to 4 that the basic results are also valid for minimum variance linear unbiased estimation and for Bayesian-type inference with or without the normality assumption. Returning to the assumption of normality, our object is to obtain the conditional distributions of α_t and α_{t+1} given Y_t for $t = 1, \dots, n$. Let $a_{t|t} = \mathrm{E}(\alpha_t|Y_t)$, $a_{t+1} = \mathrm{E}(\alpha_{t+1}|Y_t)$, $P_{t|t} = \mathrm{Var}(\alpha_t|Y_t)$ and $P_{t+1} = \mathrm{Var}(\alpha_{t+1}|Y_t)$. Since all distributions are normal, it follows from Lemma 1 that conditional distributions of subsets of variables given other subsets of variables are also normal; the distributions of α_t given Y_t and α_{t+1} given Y_t are therefore given by $\mathrm{N}(a_{t|t}, P_{t|t})$ and $\mathrm{N}(a_{t+1}, P_{t+1})$. We proceed inductively; starting with $\mathrm{N}(a_t, P_t)$, the distribution of α_t given Y_{t-1}, we show how to calculate $a_{t|t}$, a_{t+1}, $P_{t|t}$ and P_{t+1} from a_t and P_t recursively for $t = 1, \dots, n$.

Let

$$v_t = y_t - \mathrm{E}(y_t|Y_{t-1}) = y_t - \mathrm{E}(Z_t\alpha_t + \varepsilon_t|Y_{t-1}) = y_t - Z_t a_t. \tag{4.13}$$

Table 4.1 Dimensions of state space model (4.12).

Vector		**Matrix**	
y_t	$p \times 1$	Z_t	$p \times m$
α_t	$m \times 1$	T_t	$m \times m$
ε_t	$p \times 1$	H_t	$p \times p$
η_t	$r \times 1$	R_t	$m \times r$
		Q_t	$r \times r$
a_1	$m \times 1$	P_1	$m \times m$

Thus v_t is the one-step ahead forecast error of y_t given Y_{t-1}. When Y_{t-1} and v_t are fixed then Y_t is fixed and vice versa. Thus $\mathrm{E}(\alpha_t|Y_t) = \mathrm{E}(\alpha_t|Y_{t-1}, v_t)$. But $\mathrm{E}(v_t|Y_{t-1}) = \mathrm{E}(y_t - Z_t a_t|Y_{t-1}) = \mathrm{E}(Z_t\alpha_t + \varepsilon_t - Z_t a_t|Y_{t-1}) = 0$. Consequently, $\mathrm{E}(v_t) = 0$ and $\mathrm{Cov}(y_j, v_t) = \mathrm{E}[y_j \mathrm{E}(v_t|Y_{t-1})'] = 0$ for $j = 1, \dots, t-1$. Also,

$$\begin{aligned}
a_{t|t} &= \mathrm{E}(\alpha_t|Y_t) = \mathrm{E}(\alpha_t|Y_{t-1}, v_t), \\
a_{t+1} &= \mathrm{E}(\alpha_{t+1}|Y_t) = \mathrm{E}(\alpha_{t+1}|Y_{t-1}, v_t).
\end{aligned}$$

Now apply Lemma 1 in Section 4.2 to the conditional joint distribution of α_t and v_t given Y_{t-1}, taking x and y in Lemma 1 as α_t and v_t here. This gives

$$a_{t|t} = \mathrm{E}(\alpha_t|Y_{t-1}) + \mathrm{Cov}(\alpha_t, v_t)[\mathrm{Var}(v_t)]^{-1} v_t, \tag{4.14}$$

where Cov and Var refer to covariance and variance in the conditional joint distributions of α_t and v_t given Y_{t-1}. Here, $\mathrm{E}(\alpha_t|Y_{t-1}) = a_t$ by definition of a_t and

$$\begin{aligned}
\mathrm{Cov}(\alpha_t, v_t) &= \mathrm{E}\left[\alpha_t \left(Z_t\alpha_t + \varepsilon_t - Z_t a_t\right)' |Y_{t-1}\right] \\
&= \mathrm{E}\left[\alpha_t(\alpha_t - a_t)' Z_t' |Y_{t-1}\right] = P_t Z_t',
\end{aligned} \tag{4.15}$$

by definition of P_t. Let

$$F_t = \mathrm{Var}(v_t|Y_{t-1}) = \mathrm{Var}(Z_t\alpha_t + \varepsilon_t - Z_t a_t|Y_{t-1}) = Z_t P_t Z_t' + H_t. \tag{4.16}$$

Then

$$a_{t|t} = a_t + P_t Z_t' F_t^{-1} v_t. \tag{4.17}$$

By (4.3) of Lemma 1 in Section 4.2 we have

$$\begin{aligned}
P_{t|t} = \mathrm{Var}(\alpha_t|Y_t) &= \mathrm{Var}(\alpha_t|Y_{t-1}, v_t) \\
&= \mathrm{Var}(\alpha_t|Y_{t-1}) - \mathrm{Cov}(\alpha_t, v_t)[\mathrm{Var}(v_t)]^{-1} \mathrm{Cov}(\alpha_t, v_t)' \\
&= P_t - P_t Z_t' F_t^{-1} Z_t P_t.
\end{aligned} \tag{4.18}$$

We assume that F_t is nonsingular; this assumption is normally valid in well-formulated models, but in any case it is relaxed in Section 6.4. Relations (4.17) and (4.18) are sometimes called the *updating step* of the Kalman filter.

We now develop recursions for a_{t+1} and P_{t+1}. Since $\alpha_{t+1} = T_t\alpha_t + R_t\eta_t$, we have

$$\begin{aligned}
a_{t+1} &= \mathrm{E}(T_t\alpha_t + R_t\eta_t|Y_t) \\
&= T_t\,\mathrm{E}(\alpha_t|Y_t),
\end{aligned} \tag{4.19}$$

$$\begin{aligned}
P_{t+1} &= \mathrm{Var}(T_t\alpha_t + R_t\eta_t|Y_t) \\
&= T_t\ \mathrm{Var}(\alpha_t|Y_t) T_t' + R_t Q_t R_t',
\end{aligned} \tag{4.20}$$

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

Substituting (4.17) into (4.19) gives

$$\begin{aligned} a_{t+1} &= T_t a_{t|t} \\ &= T_t a_t + K_t v_t, \qquad t = 1, \ldots, n, \end{aligned} \tag{4.21}$$

where

$$K_t = T_t P_t Z_t' F_t^{-1}. \tag{4.22}$$

The matrix K_t is referred to as the *Kalman gain*. We observe that a_{t+1} has been obtained as a linear function of the previous value a_t and the forecast error v_t of y_t given Y_{t-1}. Substituting from (4.18) and (4.22) in (4.20) gives

$$P_{t+1} = T_t P_t (T_t - K_t Z_t)' + R_t Q_t R_t', \qquad t = 1, \ldots, n. \tag{4.23}$$

Relations (4.21) and (4.23) are sometimes called the *prediction step* of the Kalman filter.

The recursions (4.17), (4.21), (4.18) and (4.23) constitute the celebrated Kalman filter for model (4.12). They enable us to update our knowledge of the system each time a new observation comes in. It is noteworthy that we have derived these recursions by simple applications of the standard results of multivariate normal regression theory contained in Lemma 1. The key advantage of the recursions is that we do not have to invert a $(pt \times pt)$ matrix to fit the model each time the tth observation comes in for $t = 1, \ldots, n$; we only have to invert the $(p \times p)$ matrix F_t and p is generally much smaller than n; indeed, in the most important case in practice, $p = 1$. Although relations (4.17), (4.21), (4.18) and (4.23) constitute the forms in which the multivariate Kalman filter recursions are usually presented, we shall show in Section 6.4 that variants of them in which elements of the observational vector y_t are brought in one at a time, rather than the entire vector y_t, are in general computationally superior.

We infer from Lemma 2 that when the observations are not normally distributed and we restrict attention to estimates which are linear in y_t and unbiased, and also when matrices Z_t and T_t do not depend on previous y_t's, then under appropriate assumptions the values of $a_{t|t}$ and a_{t+1} given by the filter minimise the variance matrices of the estimates of α_t and α_{t+1} given Y_t. These considerations emphasise the point that although our results are obtained under the assumption of normality, they have a wider validity in the sense of minimum variance linear unbiased estimation when the variables involved are not normally distributed. It follows from the discussion just after the proof of Lemma 2 that the estimates are also minimum error variance linear estimates.

From the standpoint of Bayesian inference we note that, on the assumption of normality, Lemma 3 implies that the posterior densities of α_t and α_{t+1} given Y_t are normal with mean vectors (4.17) and (4.21) and variance matrices (4.18) and (4.23), respectively. We therefore do not need to provide a separate Bayesian derivation of the Kalman filter. If the assumption of normality is dropped, Lemma 4 demonstrates that the Kalman filter, as we have derived it, provides quasi-posterior mean vectors and variance matrices with minimum variance linear unbiased interpretations.

4.3.2 Kalman filter recursion

For convenience we collect together the filtering equations

$$\begin{aligned} v_t &= y_t - Z_t a_t, & F_t &= Z_t P_t Z_t' + H_t, \\ a_{t|t} &= a_t + P_t Z_t' F_t^{-1} v_t, & P_{t|t} &= P_t - P_t Z_t' F_t^{-1} Z_t P_t, \\ a_{t+1} &= T_t a_t + K_t v_t, & P_{t+1} &= T_t P_t (T_t - K_t Z_t)' + R_t Q_t R_t', \end{aligned} \tag{4.24}$$

for $t = 1, \ldots, n$, where $K_t = T_t P_t Z_t' F_t^{-1}$ with a_1 and P_1 as the mean vector and variance matrix of the initial state vector α_1. The recursion (4.24) is called the *Kalman filter*. Once $a_{t|t}$ and $P_{t|t}$ are computed, it suffices to adopt the relations

$$a_{t+1} = T_t a_{t|t}, \qquad P_{t+1} = T_t P_{t|t} T_t' + R_t Q_t R_t',$$

for predicting the state vector α_{t+1} and its variance matrix at time t. In Table 4.2 we give the dimensions of the vectors and matrices of the Kalman filter equations.

4.3.3 Kalman filter for models with mean adjustments

It is sometimes convenient to include mean adjustments in the state space model (4.12) giving the form

$$\begin{aligned} y_t &= Z_t \alpha_t + d_t + \varepsilon_t, & \varepsilon_t &\sim N(0, H_t), \\ \alpha_{t+1} &= T_t \alpha_t + c_t + R_t \eta_t, & \eta_t &\sim N(0, Q_t), \\ & & \alpha_1 &\sim N(a_1, P_1), \end{aligned} \tag{4.25}$$

where $p \times 1$ vector d_t and $m \times 1$ vector c_t are known and may change over time. Indeed, Harvey (1989) employs (4.25) as the basis for the treatment of the linear Gaussian state space model. While the simpler model (4.12) is adequate for most purposes, it is worth while presenting the Kalman filter for model (4.25) explicitly for occasional use.

Table 4.2 Dimensions of Kalman filter.

Vector		Matrix	
v_t	$p \times 1$	F_t	$p \times p$
		K_t	$m \times p$
a_t	$m \times 1$	P_t	$m \times m$
$a_{t\|t}$	$m \times 1$	$P_{t\|t}$	$m \times m$

Defining $a_t = \mathrm{E}(\alpha_t|Y_{t-1})$ and $P_t = \mathrm{Var}(\alpha_t|Y_{t-1})$ as before and assuming that d_t can depend on Y_{t-1} and c_t can depend on Y_t, the Kalman filter for (4.25) takes the form

$$\begin{aligned} v_t &= y_t - Z_t a_t - d_t, & F_t &= Z_t P_t Z_t' + H_t, \\ a_{t|t} &= a_t + P_t Z_t' F_t^{-1} v_t, & P_{t|t} &= P_t - P_t Z_t' F_t^{-1} Z_t P_t, \\ a_{t+1} &= T_t a_{t|t} + c_t, & P_{t+1} &= T_t P_{t|t} T_t' + R_t Q_t R_t', \end{aligned} \tag{4.26}$$

for $t = 1, \ldots, n$. The reader can easily verify this result by going through the argument leading from (4.19) to (4.23) step by step for model (4.25) in place of model (4.12).

4.3.4 Steady state

When dealing with a time-invariant state space model in which the system matrices Z_t, H_t, T_t, R_t, and Q_t are constant over time, the Kalman recursion for P_{t+1} converges to a constant matrix $\bar{P}$ which is the solution to the matrix equation

$$\bar{P} = T\bar{P}T' - T\bar{P}Z'\bar{F}^{-1}Z\bar{P}T' + RQR',$$

where $\bar{F} = Z\bar{P}Z' + H$. The solution that is reached after convergence to $\bar{P}$ is referred to as the *steady state solution* of the Kalman filter. Use of the steady state after convergence leads to considerable computational savings because the recursive computations for F_t, K_t, $P_{t|t}$ and P_{t+1} are no longer required.

4.3.5 State estimation errors and forecast errors

Define the *state estimation error* as

$$x_t = \alpha_t - a_t, \quad \text{with} \quad \mathrm{Var}(x_t) = P_t, \tag{4.27}$$

as for the local level model in Subsection 2.3.2. We now investigate how these errors are related to each other and to the one-step ahead forecast errors $v_t = y_t - \mathrm{E}(y_t|Y_{t-1}) = y_t - Z_t a_t$. Since v_t is the part of y_t that cannot be predicted from the past, the v_t's are sometimes referred to as *innovations*. It follows immediately from the Kalman filter relations and the definition of x_t that

$$\begin{aligned} v_t &= y_t - Z_t a_t \\ &= Z_t\alpha_t + \varepsilon_t - Z_t a_t \\ &= Z_t x_t + \varepsilon_t, \end{aligned} \tag{4.28}$$

and

$$\begin{aligned} x_{t+1} &= \alpha_{t+1} - a_{t+1} \\ &= T_t\alpha_t + R_t\eta_t - T_t a_t - K_t v_t \\ &= T_t x_t + R_t\eta_t - K_t Z_t x_t - K_t\varepsilon_t \\ &= L_t x_t + R_t\eta_t - K_t\varepsilon_t, \end{aligned} \tag{4.29}$$

where $K_t = T_t P_t Z_t' F_t^{-1}$ and $L_t = T_t - K_t Z_t$; these recursions are similar to (2.31) for the local level model. Analogously to the state space relations

$$y_t = Z_t \alpha_t + \varepsilon_t, \qquad \alpha_{t+1} = T_t \alpha_t + R_t \eta_t,$$

we obtain the *innovation analogue* of the state space model, that is,

$$v_t = Z_t x_t + \varepsilon_t, \qquad x_{t+1} = L_t x_t + R_t \eta_t - K_t \varepsilon_t, \tag{4.30}$$

with $x_1 = \alpha_1 - a_1$, for $t = 1, \dots, n$. The recursion for P_{t+1} can be derived more easily than in Subsection 4.3.1 by the steps

$$\begin{aligned} P_{t+1} &= \mathrm{Var}(x_{t+1}) = \mathrm{E}[(\alpha_{t+1} - a_{t+1})x_{t+1}'] \\ &= \mathrm{E}(\alpha_{t+1} x_{t+1}') \\ &= \mathrm{E}[(T_t \alpha_t + R_t \eta_t)(L_t x_t + R_t \eta_t - K_t \varepsilon_t)'] \\ &= T_t P_t L_t' + R_t Q_t R_t', \end{aligned}$$

since $\mathrm{Cov}(x_t, \eta_t) = 0$. Relations (4.30) will be used for deriving the smoothing recursions in the next section.

We finally show that the one-step ahead forecast errors are independent of each other using the same arguments as in Subsection 2.3.1. The joint density of the observational vectors $y_1, \dots, y_n$ is

$$p(y_1, \dots, y_n) = p(y_1) \prod_{t=2}^{n} p(y_t | Y_{t-1}).$$

Transforming from y_t to $v_t = y_t - Z_t a_t$ we have

$$p(v_1, \dots, v_n) = \prod_{t=1}^{n} p(v_t),$$

since $p(y_1) = p(v_1)$ and the Jacobian of the transformation is unity because each v_t is y_t minus a linear function of $y_1, \dots, y_{t-1}$ for $t = 2, \dots, n$. Consequently $v_1, \dots, v_n$ are independent of each other, from which it also follows that $v_t, \dots, v_n$ are independent of Y_{t-1}.

4.4 State smoothing

4.4.1 Introduction

We now derive the conditional density of α_t given the entire series $y_1, \dots, y_n$ for $t = 1, \dots, n$. We do so by assuming normality and using Lemma 1, noting from Lemmas 2 to 4 that the mean vectors and variance matrices we obtain are valid without the normality assumption in the minimum variance linear unbiased sense and are also valid for Bayesian analyses.

We shall calculate the conditional mean $\hat{\alpha}_t = \mathrm{E}(\alpha_t|Y_n)$ and the conditional variance matrix $V_t = \mathrm{Var}(\alpha_t|Y_n)$ for $t = 1, \ldots, n$. Our approach is to construct recursions for $\hat{\alpha}_t$ and V_t on the assumption that $\alpha_1 \sim \mathrm{N}(a_1, P_1)$ where a_1 and P_1 are known, deferring consideration of the case a_1 and P_1 unknown until Chapter 5. The operation of calculating $\hat{\alpha}_t$ is called *state smoothing* or just *smoothing*. A conditional mean $\mathrm{E}(\alpha_t|y_t, \ldots, y_s)$ is sometimes called a *fixed-interval smoother* to reflect the fact that it is based on the fixed time interval (t, s). The smoother conditioned on the full sample Y_n as we have just discussed is the most common smoother encountered in practice. Other types of smoothers are the *fixed-point smoother* $\hat{\alpha}_{t|n} = \mathrm{E}(\alpha_t|Y_n)$ for t fixed and $n = t+1, t+2, \ldots$ and the *fixed-lag smoother* $\hat{\alpha}_{n-j|n} = \mathrm{E}(\alpha_{n-j}|Y_n)$ for a fixed positive integer j and $n = j+1, j+2, \ldots$. We shall give formulae for these smoothers in Subsection 4.4.6. The fixed-point and fixed-lag smoothers are important in engineering; see, for example, the treatment in Chapter 7 of Anderson and Moore (1979). However, in this book we shall focus mainly on fixed-interval smoothing and when we refer simply to 'smoother' and 'smoothing', it is the fixed-interval smoother based on the full sample with which we are concerned.

4.4.2 Smoothed state vector

Take $v_1, \ldots, v_n$ as in Subsection 4.3.1 and denote the vector $(v_t', \ldots, v_n')'$ by $v_{t:n}$; note also that Y_n is fixed when Y_{t-1} and $v_{t:n}$ are fixed. To calculate $\mathrm{E}(\alpha_t|Y_n)$ and $\mathrm{Var}(\alpha_t|Y_n)$ we apply Lemma 1 of Subsection 4.2 to the conditional joint distributions of α_t and $v_{t:n}$ given Y_{t-1}, taking x and y of Lemma 1 as α_t and $v_{t:n}$ here. Using the fact that $v_t, \ldots, v_n$ are independent of Y_{t-1} and of each other with zero means, we therefore have from (4.2),

$$\begin{aligned} \hat{\alpha}_t &= \mathrm{E}(\alpha_t|Y_n) = \mathrm{E}(\alpha_t|Y_{t-1}, v_{t:n}) \\ &= a_t + \sum_{j=t}^{n} \mathrm{Cov}(\alpha_t, v_j) F_j^{-1} v_j, \end{aligned} \tag{4.31}$$

since $\mathrm{E}(\alpha_t|Y_{t-1}) = a_t$ for $t = 1, \ldots, n$, where Cov refers to covariance in the conditional distribution given Y_{t-1} and $F_j = \mathrm{Var}(v_j|Y_{t-1})$. It follows from (4.30) that

$$\begin{aligned} \mathrm{Cov}(\alpha_t, v_j) &= \mathrm{E}(\alpha_t v_j'|Y_{t-1}) \\ &= \mathrm{E}[\alpha_t(Z_j x_j + \varepsilon_j)'|Y_{t-1}] \\ &= \mathrm{E}(\alpha_t x_j'|Y_{t-1}) Z_j', \qquad j = t, \ldots, n. \end{aligned} \tag{4.32}$$

Moreover,

$$
\begin{aligned}
\mathrm{E}(\alpha_t x_t'|Y_{t-1}) &= \mathrm{E}[\alpha_t(\alpha_t - a_t)|Y_{t-1}] = P_t, \\
\mathrm{E}(\alpha_t x_{t+1}'|Y_{t-1}) &= \mathrm{E}[\alpha_t(L_t x_t + R_t \eta_t - K_t \varepsilon_t)'|Y_{t-1}] = P_t L_t', \\
\mathrm{E}(\alpha_t x_{t+2}'|Y_{t-1}) &= P_t L_t' L_{t+1}', \\
&\vdots \\
\mathrm{E}(\alpha_t x_n'|Y_{t-1}) &= P_t L_t' \cdots L_{n-1}',
\end{aligned}
\tag{4.33}
$$

using (4.30) repeatedly for $t+1, t+2, \ldots$. Note that here and elsewhere we interpret $L_t' \cdots L_{n-1}'$ as I_m when $t = n$ and as L_{n-1}' when $t = n-1$. Substituting into (4.31) gives

$$
\begin{aligned}
\hat{\alpha}_n &= a_n + P_n Z_n' F_n^{-1} v_n, \\
\hat{\alpha}_{n-1} &= a_{n-1} + P_{n-1} Z_{n-1}' F_{n-1}^{-1} v_{n-1} + P_{n-1} L_n' Z_n' F_n^{-1} v_n, \\
\hat{\alpha}_t &= a_t + P_t Z_t' F_t^{-1} v_t + P_t L_t' Z_{t+1}' F_{t+1}^{-1} v_{t+1} \\
&\quad + \cdots + P_t L_t' \cdots L_{n-1}' Z_n' F_n^{-1} v_n,
\end{aligned}
\tag{4.34}
$$

for $t = n-2, n-3, \ldots, 1$. We can therefore express the smoothed state vector as

$$
\hat{\alpha}_t = a_t + P_t r_{t-1}, \tag{4.35}
$$

where $r_{n-1} = Z_n' F_n^{-1} v_n$, $r_{n-2} = Z_{n-1}' F_{n-1}^{-1} v_{n-1} + L_{n-1}' Z_n' F_n^{-1} v_n$ and

$$
r_{t-1} = Z_t' F_t^{-1} v_t + L_t' Z_{t+1}' F_{t+1}^{-1} v_{t+1} + \cdots + L_t' L_{t+1}' \cdots L_{n-1}' Z_n' F_n^{-1} v_n, \tag{4.36}
$$

for $t = n-2, n-3, \ldots, 1$. The vector r_{t-1} is a weighted sum of innovations v_j occurring after time $t-1$, that is, for $j = t, \ldots, n$. The value at time t is

$$
r_t = Z_{t+1}' F_{t+1}^{-1} v_{t+1} + L_{t+1}' Z_{t+2}' F_{t+2}^{-1} v_{t+2} + \cdots + L_{t+1}' \cdots L_{n-1}' Z_n' F_n^{-1} v_n; \tag{4.37}
$$

also $r_n = 0$ since no innovations are available after time n. Substituting from (4.37) into (4.36) we obtain the backwards recursion

$$
r_{t-1} = Z_t' F_t^{-1} v_t + L_t' r_t, \qquad t = n, \ldots, 1, \tag{4.38}
$$

with $r_n = 0$.

Collecting these results together gives the recursion for state smoothing,

$$
\hat{\alpha}_t = a_t + P_t r_{t-1}, \qquad r_{t-1} = Z_t' F_t^{-1} v_t + L_t' r_t, \tag{4.39}
$$

for $t = n, \ldots, 1$, with $r_n = 0$; this provides an efficient algorithm for calculating $\hat{\alpha}_1, \ldots, \hat{\alpha}_n$. This smoother, together with the recursion for computing the variance matrix of the smoothed state vector which we present in Subsection 4.4.3, was proposed in the forms (4.39) and (4.43) below by de Jong (1988a), de Jong (1989) and Kohn and Ansley (1989) although the earlier treatments in the engineering literature by Bryson and Ho (1969) and Young (1984) are similar.

4.4.3 Smoothed state variance matrix

A recursion for calculating $V_t = \mathrm{Var}(\alpha_t|Y_n)$ will now be derived. We have defined $v_{t:n} = (v_t', \dots, v_n')'$. Applying Lemma 1 of Section 4.2 to the conditional joint distribution of α_t and $v_{t:n}$ given Y_{t-1}, taking $x = \alpha_t$ and $y = v_{t:n}$, we obtain from (4.3)

$$V_t = \mathrm{Var}(\alpha_t|Y_{t-1}, v_{t:n}) = P_t - \sum_{j=t}^{n} \mathrm{Cov}(\alpha_t, v_j) F_j^{-1} \mathrm{Cov}(\alpha_t, v_j)',$$

where $\mathrm{Cov}(\alpha_t, vj)$ and F_j are as in (4.31), since $v_t, \dots, v_n$ are independent of each other and of Y_{t-1} with zero means. Using (4.32) and (4.33) we obtain immediately

$$\begin{aligned} V_t &= P_t - P_t Z_t' F_t^{-1} Z_t P_t - P_t L_t' Z_{t+1}' F_{t+1}^{-1} Z_{t+1} L_t P_t - \cdots \\ &\quad - P_t L_t' \cdots L_{n-1}' Z_n' F_n^{-1} Z_n L_{n-1} \cdots L_t P_t \\ &= P_t - P_t N_{t-1} P_t, \end{aligned}$$

where

$$\begin{aligned} N_{t-1} &= Z_t' F_t^{-1} Z_t + L_t' Z_{t+1}' F_{t+1}^{-1} Z_{t+1} L_t + \cdots \\ &\quad + L_t' \cdots L_{n-1}' Z_n' F_n^{-1} Z_n L_{n-1} \cdots L_t. \end{aligned} \tag{4.40}$$

We note that here, as in the previous subsection, we interpret $L_t' \cdots L_{n-1}'$ as I_m when $t = n$ and as L_{n-1}' when $t = n - 1$. The value at time t is

$$\begin{aligned} N_t &= Z_{t+1}' F_{t+1}^{-1} Z_{t+1} + L_{t+1}' Z_{t+2}' F_{t+2}^{-1} Z_{t+2} L_{t+1} + \cdots \\ &\quad + L_{t+1}' \cdots L_{n-1}' Z_n' F_n^{-1} Z_n L_{n-1} \cdots L_{t+1}. \end{aligned} \tag{4.41}$$

Substituting (4.41) into (4.40) we obtain the backwards recursion

$$N_{t-1} = Z_t' F_t^{-1} Z_t + L_t' N_t L_t, \qquad t = n, \dots, 1. \tag{4.42}$$

Noting from (4.41) that $N_{n-1} = Z_n' F_n^{-1} Z_n$ we deduce that recursion (4.42) is initialised with $N_n = 0$. Collecting these results, we find that V_t can be efficiently calculated by the recursion

$$N_{t-1} = Z_t' F_t^{-1} Z_t + L_t' N_t L_t, \qquad V_t = P_t - P_t N_{t-1} P_t, \tag{4.43}$$

for $t = n, \dots, 1$ with $N_n = 0$. Since $v_{t+1}, \dots, v_n$ are independent it follows from (4.37) and (4.41) that $N_t = \mathrm{Var}(r_t)$.

4.4.4 State smoothing recursion

For convenience we collect together the smoothing equations for the state vector,

$$
\begin{aligned}
r_{t-1} &= Z_t' F_t^{-1} v_t + L_t' r_t, & N_{t-1} &= Z_t' F_t^{-1} Z_t + L_t' N_t L_t, \\
\hat{\alpha}_t &= a_t + P_t r_{t-1}, & V_t &= P_t - P_t N_{t-1} P_t,
\end{aligned}
\tag{4.44}
$$

for $t = n, \ldots, 1$ initialised with $r_n = 0$ and $N_n = 0$. We refer to these collectively as the *state smoothing recursion*. As noted earlier, Lemmas 2 to 4 imply that the recursion (4.44) is also valid for non-normal cases in the MVLUE sense and for Bayesian analyses. Taken together, the recursions (4.24) and (4.44) will be referred to as the *Kalman filter and smoother*. We see that the way the filtering and smoothing is performed is that we proceed forwards through the series using (4.24) and backwards through the series using (4.44) to obtain $\hat{\alpha}_t$ and V_t for $t = 1, \ldots, n$. During the forward pass we need to store the quantities v_t, F_t, K_t, a_t and P_t for $t = 1, \ldots, n$. Alternatively we can store a_t and P_t only and recalculate v_t, F_t and K_t using a_t and P_t but this is usually not done since the dimensions of v_t, F_t and K_t are usually small relative to a_t and P_t, so the extra storage required is small. In Table 4.3 we present the dimensions of the vectors and matrices of the smoothing equations of this section and Subsection 4.5.3.

4.4.5 Updating smoothed estimates

In many situations observations come in one at a time and we wish to update the smoothed estimates each time a new observation comes in. We shall develop recursions for doing this which are computationally more efficient than applying (4.44) repeatedly.

Denote the new observation by y_{n+1} and suppose that we wish to calculate $\hat{\alpha}_{t|n+1} = \mathrm{E}(\alpha_t|Y_{n+1})$. For convenience, we relabel $\hat{\alpha}_t = \mathrm{E}(\alpha_t|Y_n)$ as $\hat{\alpha}_{t|n}$. We have

$$
\begin{aligned}
\hat{\alpha}_{t|n+1} &= \mathrm{E}(\alpha_t|Y_n, v_{n+1}) \\
&= \hat{\alpha}_{t|n} + \mathrm{Cov}(\alpha_t, v_{n+1}) F_{n+1}^{-1} v_{n+1},
\end{aligned}
$$

Table 4.3 Dimensions of smoothing recursions of Subsections 4.4.4 and 4.5.3.

Vector		Matrix	
r_t	$m \times 1$	N_t	$m \times m$
$\hat{\alpha}_t$	$m \times 1$	V_t	$m \times m$
u_t	$p \times 1$	D_t	$p \times p$
$\hat{\varepsilon}_t$	$p \times 1$		
$\hat{\eta}_t$	$r \times 1$		

by Lemma 1. From (4.32) and (4.33) we have

$$\mathrm{Cov}(\alpha_t, v_{n+1}) = P_t L_t' \dots L_n' Z_{n+1}',$$

giving

$$\hat{\alpha}_{t|n+1} = \hat{\alpha}_{t|n} + P_t L_t' \dots L_n' Z_{n+1}' F_{n+1}^{-1} v_{n+1}, \tag{4.45}$$

for $t = 1, \dots, n$. In addition, from (4.17),

$$\hat{\alpha}_{n+1|n+1} = a_{n+1} + P_{n+1} Z_{n+1}' F_{n+1}^{-1} v_{n+1}. \tag{4.46}$$

Now consider the updating of the smoothed state variance matrix $V_t = \mathrm{Var}(\alpha_t|Y_n)$. For convenience we relabel this as $V_{t|n}$. Let $V_{t|n+1} = \mathrm{Var}(\alpha_t|Y_{n+1})$. By Lemma 1,

$$\begin{aligned}
V_{t|n+1} &= \mathrm{Var}(\alpha_t|Y_n, v_{n+1}) \\
&= \mathrm{Var}(\alpha_t|Y_n) - \mathrm{Cov}(\alpha_t, v_{n+1}) F_{n+1}^{-1} \mathrm{Cov}(\alpha_t, v_{n+1})' \\
&= V_{t|n} - P_t L_t' \dots L_n' Z_{n+1}' F_{n+1}^{-1} Z_{n+1} L_n \dots L_t P_t,
\end{aligned} \tag{4.47}$$

for $t = 1, \dots, n$ with

$$V_{n+1|n+1} = P_{n+1} - P_{n+1} Z_{n+1}' F_{n+1}^{-1} Z_{n+1} P_{n+1}. \tag{4.48}$$

Let $b_{t|n+1} = L_t' \dots L_n'$ with $b_{n+1|n+1} = I_m$. Then $b_{t|n+1} = L_t' b_{t+1|n+1}$ for $t = n, \dots, 1$ and we can write recursions (4.45) and (4.47) in the compact forms

$$\hat{\alpha}_{t|n+1} = \hat{\alpha}_{t|n} + P_t b_{t|n+1} Z_{n+1}' F_{n+1}^{-1} v_{n+1}, \tag{4.49}$$

$$V_{t|n+1} = V_{t|n} - P_t b_{t|n+1} Z_{n+1}' F_{n+1}^{-1} Z_{n+1} b_{t|n+1}' P_t, \tag{4.50}$$

for $n = t, t+1, \dots$ with $\hat{\alpha}_{n|n} = a_n + P_n Z_n' F_n^{-1} v_n$ and $V_{n|n} = P_n - P_n Z_n' F_n^{-1} Z_n P_n$. Note that P_t, L_t, F_{n+1} and v_{n+1} are all readily available from the Kalman filter.

4.4.6 Fixed-point and fixed-lag smoothers

The fixed-point smoother $\hat{\alpha}_{t|n} = \mathrm{E}(\alpha_t|Y_n)$ for t fixed and $n = t+1, t+2, \dots$ is given directly by the recursion (4.49) and its error variance matrix is given by (4.50).

From (4.39) the fixed-lag smoother $\hat{\alpha}_{n-j|n} = \mathrm{E}(\alpha_{n-j}|Y_n)$ for j fixed at possible values $0, 1, \dots, n-1$ and $n = j+1, j+2, \dots$ is given by

$$\hat{\alpha}_{n-j|n} = a_{n-j} + P_{n-j} r_{n-j-1}, \tag{4.51}$$

where r_{n-j-1} is obtained from the backward recursion

$$r_{t-1} = Z_t' F_t^{-1} v_t + L_t' r_t, \qquad t = n, \dots, n-j, \tag{4.52}$$

with $r_n = 0$. From (4.43) its error variance matrix is given by

$$V_{n-j|n} = P_{n-j} - P_{n-j}N_{n-j-1}P_{n-j}, \tag{4.53}$$

where N_{n-j-1} is obtained from the backward recursion

$$N_{t-1} = Z_t'F_t^{-1}Z_t + L_t'N_tL_t, \qquad t = n, \dots, n-j, \tag{4.54}$$

with $N_n = 0$, for $j = 0, 1, \dots, n-1$.

4.5 Disturbance smoothing

In this section we will derive recursions for computing the smoothed estimates $\hat{\varepsilon}_t = \mathrm{E}(\varepsilon_t|Y_n)$ and $\hat{\eta}_t = \mathrm{E}(\eta_t|Y_n)$ of the disturbance vectors ε_t and η_t given all the observations $y_1, \dots, y_n$. These estimates have a variety of uses, particularly for parameter estimation and diagnostic checking, as will be indicated in Sections 7.3 and 7.5.

4.5.1 Smoothed disturbances

Let $\hat{\varepsilon}_t = \mathrm{E}(\varepsilon_t|Y_n)$. By Lemma 1 we have

$$\hat{\varepsilon}_t = \mathrm{E}(\varepsilon_t|Y_{t-1}, v_t, \dots, v_n) = \sum_{j=t}^{n} \mathrm{E}(\varepsilon_t v_j')F_j^{-1}v_j, \qquad t = 1, \dots, n, \tag{4.55}$$

since $\mathrm{E}(\varepsilon_t|Y_{t-1}) = 0$ and ε_t and v_t are jointly independent of Y_{t-1}. It follows from (4.30) that $\mathrm{E}(\varepsilon_t v_j') = \mathrm{E}(\varepsilon_t x_j')Z_j' + \mathrm{E}(\varepsilon_t \varepsilon_j')$ with $\mathrm{E}(\varepsilon_t x_t') = 0$ for $t = 1, \dots, n$ and $j = t, \dots, n$. Therefore

$$\mathrm{E}(\varepsilon_t v_j') = \begin{cases} H_t, & j = t, \\ \mathrm{E}(\varepsilon_t x_j')Z_j', & j = t+1, \dots, n, \end{cases} \tag{4.56}$$

with

$$\begin{aligned} \mathrm{E}(\varepsilon_t x_{t+1}') &= -H_tK_t', \\ \mathrm{E}(\varepsilon_t x_{t+2}') &= -H_tK_t'L_{t+1}', \\ &\vdots \\ \mathrm{E}(\varepsilon_t x_n') &= -H_tK_t'L_{t+1}' \cdots L_{n-1}', \end{aligned} \tag{4.57}$$

which follow from (4.30), for $t = 1, \dots, n-1$. Note that here as elsewhere we interpret $L_{t+1}' \cdots L_{n-1}'$ as I_m when $t = n-1$ and as L_{n-1}' when $t = n-2$. Substituting (4.56) into (4.55) leads to

$$
\begin{aligned}
\hat{\varepsilon}_t &= H_t(F_t^{-1}v_t - K_t'Z_{t+1}'F_{t+1}^{-1}v_{t+1} - K_t'L_{t+1}'Z_{t+2}'F_{t+2}^{-1}v_{t+2} - \cdots \\
&\qquad - K_t'L_{t+1}'\cdots L_{n-1}'Z_n'F_n^{-1}v_n) \\
&= H_t(F_t^{-1}v_t - K_t'r_t) \\
&= H_t u_t, \qquad t = n,\ldots,1,
\end{aligned}
\tag{4.58}
$$

where r_t is defined in (4.37) and

$$
u_t = F_t^{-1}v_t - K_t'r_t. \tag{4.59}
$$

We refer to the vector u_t as the *smoothing error*.

The smoothed estimate of η_t is denoted by $\hat{\eta}_t = \mathrm{E}(\eta_t|Y_n)$ and analogously to (4.55) we have

$$
\hat{\eta}_t = \sum_{j=t}^{n} \mathrm{E}(\eta_t v_j')F_j^{-1}v_j, \qquad t = 1,\ldots,n. \tag{4.60}
$$

The relations (4.30) imply that

$$
\mathrm{E}(\eta_t v_j') = \begin{cases} Q_t R_t' Z_{t+1}', & j = t+1, \\ \mathrm{E}(\eta_t x_j')Z_j', & j = t+2,\ldots,n, \end{cases} \tag{4.61}
$$

with

$$
\begin{aligned}
\mathrm{E}(\eta_t x_{t+2}') &= Q_t R_t' L_{t+1}', \\
\mathrm{E}(\eta_t x_{t+3}') &= Q_t R_t' L_{t+1}' L_{t+2}', \\
&\ \,\vdots \\
\mathrm{E}(\eta_t x_n') &= Q_t R_t' L_{t+1}' \cdots L_{n-1}',
\end{aligned}
\tag{4.62}
$$

for $t = 1,\ldots,n-1$. Substituting (4.61) into (4.60) and noting that $\mathrm{E}(\eta_t v_t') = 0$ leads to

$$
\begin{aligned}
\hat{\eta}_t &= Q_t R_t' \left(Z_{t+1}'F_{t+1}^{-1}v_{t+1} + L_{t+1}'Z_{t+2}'F_{t+2}^{-1}v_{t+2} + \cdots + L_{t+1}'\cdots L_{n-1}'Z_n'F_n^{-1}v_n \right) \\
&= Q_t R_t' r_t, \qquad t = n,\ldots,1,
\end{aligned}
\tag{4.63}
$$

where r_t is obtained from (4.38). This result is useful as we will show in the next section but it also gives the vector r_t the interpretation as the ‘scaled’ smoothed estimator of η_t. Note that in many practical cases the matrix $Q_t R_t'$ is diagonal or sparse. Equations (4.58) and (4.65) below were first given by de Jong (1988a) and Kohn and Ansley (1989). Equations (4.63) and (4.68) below were given by Koopman (1993). Earlier developments on disturbance smoothing have been reported by Kailath and Frost (1968).

4.5.2 Smoothed disturbance variance matrices

The error variance matrices of the smoothed disturbances are developed by the same approach that we used in Subsection 4.4.3 to derive the smoothed state variance matrix. Using Lemma 1 we have

$$\begin{aligned}\mathrm{Var}(\varepsilon_t|Y_n) &= \mathrm{Var}(\varepsilon_t|Y_{t-1}, v_t, \dots, v_n)\\ &= \mathrm{Var}(\varepsilon_t|Y_{t-1}) - \sum_{j=t}^{n} \mathrm{Cov}(\varepsilon_t, v_j)\,\mathrm{Var}(v_j)^{-1}\,\mathrm{Cov}(\varepsilon_t, v_j)'\\ &= H_t - \sum_{j=t}^{n} \mathrm{Cov}(\varepsilon_t, v_j) F_j^{-1}\,\mathrm{Cov}(\varepsilon_t, v_j)', \end{aligned} \tag{4.64}$$

where $\mathrm{Cov}(\varepsilon_t, v_j) = \mathrm{E}(\varepsilon_t v_j')$ which is given by (4.56). By substitution we obtain

$$\begin{aligned}\mathrm{Var}(\varepsilon_t|Y_n) &= H_t - H_t \big(F_t^{-1} + K_t' Z_{t+1}' F_{t+1}^{-1} Z_{t+1} K_t\\ &\quad + K_t' L_{t+1}' Z_{t+2}' F_{t+2}^{-1} Z_{t+2} L_{t+1} K_t + \cdots\\ &\quad + K_t' L_{t+1}' \cdots L_{n-1}' Z_n' F_n^{-1} Z_n L_{n-1} \cdots L_{t+1} K_t\big) H_t'\\ &= H_t - H_t \left(F_t^{-1} + K_t' N_t K_t\right) H_t\\ &= H_t - H_t D_t H_t, \end{aligned} \tag{4.65}$$

with

$$D_t = F_t^{-1} + K_t' N_t K_t, \tag{4.66}$$

where N_t is defined in (4.41) and can be obtained from the backward recursion (4.42).

In a similar way the variance matrix $\mathrm{Var}(\eta_t|Y_n)$ is given by

$$\mathrm{Var}(\eta_t|Y_n) = \mathrm{Var}(\eta_t) - \sum_{j=t}^{n} \mathrm{Cov}(\eta_t, v_j) F_j^{-1}\,\mathrm{Cov}(\eta_t, v_j)', \tag{4.67}$$

where $\mathrm{Cov}(\eta_t, v_j) = \mathrm{E}(\eta_t, v_j')$ which is given by (4.61). Substitution gives

$$\begin{aligned}\mathrm{Var}(\eta_t|Y_n) &= Q_t - Q_t R_t' \big(Z_{t+1}' F_{t+1}^{-1} Z_{t+1} + L_{t+1}' Z_{t+2}' F_{t+2}^{-1} Z_{t+2} L_{t+1} + \cdots\\ &\quad + L_{t+1}' \cdots L_{n-1}' Z_n' F_n^{-1} Z_n L_{n-1} \cdots L_{t+1}\big) R_t Q_t\\ &= Q_t - Q_t R_t' N_t R_t Q_t, \end{aligned} \tag{4.68}$$

where N_t is obtained from (4.42).

4.5.3 Disturbance smoothing recursion

For convenience we collect together the smoothing equations for the disturbance vectors,

$$
\begin{aligned}
\hat{\varepsilon}_t &= H_t\left(F_t^{-1}v_t - K_t'r_t\right), & \mathrm{Var}(\varepsilon_t|Y_n) &= H_t - H_t\left(F_t^{-1} + K_t'N_tK_t\right)H_t,\\
\hat{\eta}_t &= Q_tR_t'r_t, & \mathrm{Var}(\eta_t|Y_n) &= Q_t - Q_tR_t'N_tR_tQ_t,\\
r_{t-1} &= Z_t'F_t^{-1}v_t + L_t'r_t, & N_{t-1} &= Z_t'F_t^{-1}Z_t + L_t'N_tL_t,
\end{aligned}
\tag{4.69}
$$

for $t = n, \ldots, 1$ where $r_n = 0$ and $N_n = 0$. These equations can be reformulated as

$$
\begin{aligned}
\hat{\varepsilon}_t &= H_tu_t, & \mathrm{Var}(\varepsilon_t|Y_n) &= H_t - H_tD_tH_t,\\
\hat{\eta}_t &= Q_tR_t'r_t, & \mathrm{Var}(\eta_t|Y_n) &= Q_t - Q_tR_t'N_tR_tQ_t,\\
u_t &= F_t^{-1}v_t - K_t'r_t, & D_t &= F_t^{-1} + K_t'N_tK_t,\\
r_{t-1} &= Z_t'u_t + T_t'r_t, & N_{t-1} &= Z_t'D_tZ_t + T_t'N_tT_t - Z_t'K_t'N_tT_t - T_t'N_tK_tZ_t,
\end{aligned}
$$

for $t = n, \ldots, 1$, which are computationally more efficient since they rely directly on the system matrices Z_t and T_t which have the property that they usually contain many zeros and ones. We refer to these equations collectively as the *disturbance smoothing recursion*. The smoothing error u_t and vector r_t are important in their own right for a variety of reasons which we will discuss in Section 7.5. The dimensions of the vectors and matrices of disturbance smoothing are given in Table 4.3.

We have shown that disturbance smoothing is performed in a similar way to state smoothing: we proceed forward through the series using (4.24) and backward through the series using (4.69) to obtain $\hat{\varepsilon}_t$ and $\hat{\eta}_t$ together with the corresponding conditional variances for $t = 1, \ldots, n$. The storage requirement for (4.69) during the forward pass is less than for the state smoothing recursion (4.44) since here we only need v_t, F_t and K_t of the Kalman filter. Also, the computations are quicker for disturbance smoothing since they do not involve the vector a_t and the matrix P_t which are not sparse.

4.6 Other state smoothing algorithms

4.6.1 Classical state smoothing

Alternative algorithms for state smoothing have also been proposed. For example, Anderson and Moore (1979) present the so-called *classical fixed-interval smoother*, due to Rauch, Tung and Striebel (1965), which for our state space model is given by

$$
\hat{\alpha}_t = a_{t|t} + P_{t|t}T_t'P_{t+1}^{-1}(\hat{\alpha}_{t+1} - a_{t+1}), \qquad t = n, \ldots, 1, \tag{4.70}
$$

where

$$a_{t|t} = \mathrm{E}(\alpha_t|Y_t) = a_t + P_t Z_t' F_t^{-1} v_t, \qquad P_{t|t} = \mathrm{Var}(\alpha_t|Y_t) = P_t - P_t Z_t' F_t^{-1} Z_t P_t;$$

see equations (4.17) and (4.18). Notice that $T_t P_{t|t} = L_t P_t$.

Following Koopman (1998), we now show that (4.39) can be derived from (4.70). Substituting for $a_{t|t}$ and $T_t P_{t|t}$ into (4.70) we have

$$\hat{\alpha}_t = a_t + P_t Z_t' F_t^{-1} v_t + P_t L_t' P_{t+1}^{-1}(\hat{\alpha}_{t+1} - a_{t+1}).$$

By defining $r_t = P_{t+1}^{-1}(\hat{\alpha}_{t+1} - a_{t+1})$ and re-ordering the terms, we obtain

$$P_t^{-1}(\hat{\alpha}_t - a_t) = Z_t' F_t^{-1} v_t + L_t' P_{t+1}^{-1}(\hat{\alpha}_{t+1} - a_{t+1}),$$

and hence

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

which is (4.38). Note that the alternative definition of r_t also implies that $r_n = 0$. Finally, it follows immediately from the definitional relation $r_{t-1} = P_t^{-1}(\hat{\alpha}_t - a_t)$ that $\hat{\alpha}_t = a_t + P_t r_{t-1}$.

A comparison of the two different algorithms shows that the Anderson and Moore smoother requires inversion of $n-1$ possibly large matrices P_t whereas the smoother (4.39) requires no inversion other than of F_t which has been inverted as part of the computations of the Kalman filter. This is a considerable advantage for large models. For both smoothers the Kalman filter vector a_t and matrix P_t need to be stored together with v_t, F_t^{-1} and K_t, for $t = 1, \dots, n$. The state smoothing equation of Koopman (1993), which we consider in Subsection 4.6.2, does not involve a_t and P_t and it therefore leads to further computational savings.

4.6.2 Fast state smoothing

The smoothing recursion for the disturbance vector η_t in Subsection 4.5.3 is particularly useful since it leads to a computationally more efficient method of calculating $\hat{\alpha}_t$ for $t = 1, \dots, n$ than (4.39). Given the state equation

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t,$$

it follows by taking expectations given Y_n that

$$\begin{aligned} \hat{\alpha}_{t+1} &= T_t \hat{\alpha}_t + R_t \hat{\eta}_t \\ &= T_t \hat{\alpha}_t + R_t Q_t R_t' r_t, \qquad t = 1, \dots, n, \end{aligned} \tag{4.71}$$

which is initialised via the relation (4.35) for $t = 1$, that is, $\hat{\alpha}_1 = a_1 + P_1 r_0$ where r_0 is obtained from (4.38). This recursion, due to Koopman (1993), can be used to generate the smoothed states $\hat{\alpha}_1, \ldots, \hat{\alpha}_n$ by an algorithm different from (4.39); it does not require the storage of a_t and P_t and it does not involve multiplications by the full matrix P_t, for $t = 1, \ldots, n$. After the Kalman filter and the storage of v_t, F_t^{-1} and K_t has taken place, the backwards recursion (4.38) is undertaken and the vector r_t is stored for which the storage space of K_t can be used so no additional storage space is required. It should be kept in mind that the matrices T_t and $R_t Q_t R_t'$ are usually sparse matrices containing many zero and unity values which make the application of (4.71) rapid; this property does not apply to P_t which is a full variance matrix. This approach, however, cannot be used to obtain a recursion for the calculation of $V_t = \mathrm{Var}(\alpha_t | Y_n)$; if V_t is required then (4.39) and (4.43) should be used.

4.6.3 The Whittle relation between smoothed estimates

Whittle (1991) has provided an interesting relationship between smoothed state vector estimates based on direct likelihood arguments. In particular, for the local level model (2.3) this relationship reduces to

$$\hat{\alpha}_{t-1} = 2\hat{\alpha}_t - \hat{\alpha}_{t+1} - q\,(y_t - \hat{\alpha}_t)\,,$$

for $t = n, n-1, \ldots, 1$ with $\hat{\alpha}_{n+1} = \hat{\alpha}_n$. The initialisation and recursive equations follow directly from the loglikelihood function written as a joint density for $\alpha_1, \ldots, \alpha_n, \alpha_{n+1}$ and $y_1, \ldots, y_n$. Taking first derivatives with respect to α_t and solving the equations when set to zero yields the result. The results also hold for the general model. The recursive algorithm is appealing since the storage of Kalman filter quantities is not needed; the computation of $\hat{\alpha}_{n+1} = a_{n+1|n}$ is only required by the Kalman filter. However the algorithm is numerically unstable due to the accumulation of numerical inaccuracies.

4.6.4 Two filter formula for smoothing

Mayne (1966), Fraser and Potter (1969) and Kitagawa (1994) have developed a method of smoothing based on reversing the observations and the state space model equations (4.12). Reversion implies that the observation sequence $y_1, \ldots, y_n$ becomes

$$y_1^-, \ldots, y_n^- \;\equiv\; y_n, \ldots, y_1.$$

We define the vector $Y_t^- = \left(y_1^{-\prime}, \ldots, y_t^{-\prime}\right)'$ for $t = 1, \ldots, n$. Here, Y_1^- is the $p \times 1$ vector y_n while Y_t and Y_t^- are two vectors which have different number of elements in almost all cases. The vectors Y_n and Y_n^- consist of the same elements but the elements are ordered in opposite directions. It is suggested that the same state space model can be considered for the time series in reverse order, that is

$$y_t^- = Z_t^- \alpha_t^- + \varepsilon_t^-, \qquad \alpha_{t+1}^- = T_t^- \alpha_t^- + R_t^- \eta_t^-, \tag{4.72}$$

for $t = 1, \ldots, n$ where x_t^- is the t-th last element of the sequence $x_1, \ldots, x_n$ for any variable x. It seems hard to justify this approach generally. However, in the case that the initial state vector is fully diffuse, that is $\alpha_1 \sim \mathrm{N}(0, \kappa I)$ with $\kappa \to \infty$, it can be justified. We illustrate this for the local level model of Chapter 2 using its matrix representation.

When the $n \times 1$ time series vector Y_n is generated by the univariate local level model, we have from (2.4) that $Y_n \sim \mathrm{N}(a_1 \mathbf{1}, \Omega)$ where a_1 is the initial state and Ω is specified as in (2.4) and (2.5). Let J here be the $n \times n$ matrix with its $(i, n-i+1)$ element equal to unity for $i = 1, \dots, n$ and with other elements equal to zero. Vector JY_n is then equal to Y_n in reverse order, that is Y_n^-. Matrix J has the properties $J' = J^{-1} = J$ and $J\,J = I$. In Subsection 2.3.1 it is argued that the Kalman filter effectively carries out the computation $Y_n' \Omega^{-1} Y_n = v' F^{-1} v$ via the Cholesky decomposition. When this computation can also be done in reverse order, it is implied that

$$Y_n' \Omega^{-1} Y_n = (JY_n)' \, \Omega^{-1} J Y_n = Y_n' J \Omega^{-1} J Y_n.$$

Hence we have implied that $\Omega^{-1} = J\Omega^{-1}J$. This property does not hold for a symmetric matrix generally. However, if Ω^{-1} is a symmetric Toeplitz matrix, the property holds. When matrix Ω is defined as (2.4), that is $\Omega = \mathbf{1}\mathbf{1}' P_1 + \Sigma$ where P_1 is the initial state variance and Σ is defined in (2.5), and when $P_1 = \kappa \to \infty$, Ω^{-1} converges to a symmetric Toeplitz matrix and hence the property holds. We can therefore reverse the ordering of the local level observations and obtain the same value for $v' F^{-1} v$. This argument can be generalised for state space models where the state vector is fully diffuse. In other cases, the justification of reversing the observations is harder to establish.

When reversing the observations is valid, the following two filtering methods for smoothing can be applied. The smoothed density of the state vector can be expressed by

$$\begin{aligned} p(\alpha_t | Y_n) &= p(\alpha_t | Y_{t-1}, Y_t^-) \\ &= c\, p(\alpha_t, Y_t^- | Y_{t-1}) \\ &= c\, p(\alpha_t | Y_{t-1})\, p(Y_t^- | \alpha_t, Y_{t-1}) \\ &= c\, p(\alpha_t | Y_{t-1})\, p(Y_t^- | \alpha_t), \end{aligned} \tag{4.73}$$

where c is some constant that does not depend on α_t. The last equality (4.73) holds since the disturbances $\varepsilon_t, \dots, \varepsilon_n$ and $\eta_t, \dots, \eta_{n-1}$ of (4.12) do not depend on Y_{t-1}. Taking logs of (4.73), multiplying it by -2 and only writing the terms associated with α_t, gives the equality

$$(\alpha_t - a_{t|n})' P_{t|n}^{-1} (\alpha_t - a_{t|n}) = (\alpha_t - a_t)' P_t^{-1} (\alpha_t - a_t) + (\alpha_t - a_{t|t}^-)' Q_{t|t}^{-1} (\alpha_t - a_{t|t}^-),$$

where

$$a_{t|t}^- = \mathrm{E}(\alpha_t | Y_t^-) = \mathrm{E}(\alpha_t | y_t, \dots, y_n), \qquad Q_{t|t} = \mathrm{Var}(\alpha_t | Y_t^-) = \mathrm{Var}(\alpha_t | y_t, \dots, y_n).$$

Vector $a^-_{t|t}$, associated with $a_{t|t}$, and matrix $Q_{t|t}$, associated with $P_{t|t}$, are obtained from the Kalman filter applied to the observations in reverse order, that is, $y^-_1, \dots, y^-_n$, and based on the model (4.72). By some minor matrix manipulations of the above equality, we obtain

$$a_{t|n} = P_{t|n}(P_t^{-1}a_t + Q_{t|t}^{-1}a^-_{t|t}), \qquad P_{t|n} = \left(P_t^{-1} + Q_{t|t}^{-1}\right)^{-1}, \tag{4.74}$$

for $t = 1, \dots, n$. To avoid taking inverses of the variance matrices, the application of an information filter can be considered; see Anderson and Moore (1979, Chapter 3) for a discussion of the information filter.

4.7 Covariance matrices of smoothed estimators

In this section we develop expressions for the covariances between the errors of the smoothed estimators $\hat{\varepsilon}_t$, $\hat{\eta}_t$ and $\hat{\alpha}_t$ contemporaneously and for all leads and lags.

It turns out that the covariances of smoothed estimators rely basically on the cross-expectations $\mathrm{E}(\varepsilon_t r_j')$, $\mathrm{E}(\eta_t r_j')$ and $\mathrm{E}(\alpha_t r_j')$ for $j = t+1, \dots, n$. To develop these expressions we collect from equations (4.56), (4.57), (4.61), (4.62), (4.33) and (4.32) the results

$$\begin{aligned}
\mathrm{E}(\varepsilon_t x_t') &= 0, & \mathrm{E}(\varepsilon_t v_t') &= H_t,\\
\mathrm{E}(\varepsilon_t x_j') &= -H_t K_t' L_{t+1}' \cdots L_{j-1}', & \mathrm{E}(\varepsilon_t v_j') &= \mathrm{E}(\varepsilon_t x_j') Z_j',\\
\mathrm{E}(\eta_t x_t') &= 0, & \mathrm{E}(\eta_t v_t') &= 0,\\
\mathrm{E}(\eta_t x_j') &= Q_t R_t' L_{t+1}' \cdots L_{j-1}', & \mathrm{E}(\eta_t v_j') &= \mathrm{E}(\eta_t x_j') Z_j',\\
\mathrm{E}(\alpha_t x_t') &= P_t, & \mathrm{E}(\alpha_t v_t') &= P_t Z_t',\\
\mathrm{E}(\alpha_t x_j') &= P_t L_t' L_{t+1}' \cdots L_{j-1}', & \mathrm{E}(\alpha_t v_j') &= \mathrm{E}(\alpha_t x_j') Z_j',
\end{aligned} \tag{4.75}$$

for $j = t+1, \dots, n$. For the case $j = t+1$, we replace $L_{t+1}' \cdots L_t'$ by the identity matrix I.

We derive the cross-expectations below using the definitions

$$\begin{aligned}
r_j &= \sum_{k=j+1}^{n} L_{j+1}' \cdots L_{k-1}' Z_k' F_k^{-1} v_k,\\
N_j &= \sum_{k=j+1}^{n} L_{j+1}' \cdots L_{k-1}' Z_k' F_k^{-1} Z_k L_{k-1} \cdots L_{j+1},
\end{aligned}$$

which are given by (4.36) and (4.40), respectively. It follows that

$$\begin{aligned}
\mathrm{E}(\varepsilon_t r_j') &= \mathrm{E}(\varepsilon_t v_{j+1}')F_{j+1}^{-1}Z_{j+1} + \mathrm{E}(\varepsilon_t v_{j+2}')F_{j+2}^{-1}Z_{j+2}L_{j+1} + \cdots \\
&\quad + \mathrm{E}(\varepsilon_t v_n')F_n^{-1}Z_n L_{n-1}\cdots L_{j+1} \\
&= -H_t K_t' L_{t+1}'\cdots L_j' Z_{j+1}' F_{j+1}^{-1} Z_{j+1} \\
&\quad - H_t K_t' L_{t+1}'\cdots L_{j+1}' Z_{j+2}' F_{j+2}^{-1} Z_{j+2} L_{j+1} - \cdots \\
&\quad - H_t K_t' L_{t+1}'\cdots L_{n-1}' Z_n' F_n^{-1} Z_n L_{n-1}\cdots L_{j+1} \\
&= -H_t K_t' L_{t+1}'\cdots L_{j-1}' L_j' N_j, \qquad (4.76)
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}(\eta_t r_j') &= \mathrm{E}(\eta_t v_{j+1}')F_{j+1}^{-1}Z_{j+1} + \mathrm{E}(\eta_t v_{j+2}')F_{j+2}^{-1}Z_{j+2}L_{j+1} + \cdots \\
&\quad + \mathrm{E}(\eta_t v_n')F_n^{-1}Z_n L_{n-1}\cdots L_{j+1} \\
&= Q_t R_t' L_{t+1}'\cdots L_j' Z_{j+1}' F_{j+1}^{-1} Z_{j+1} \\
&\quad + Q_t R_t' L_{t+1}'\cdots L_{j+1}' Z_{j+2}' F_{j+2}^{-1} Z_{j+2} L_{j+1} + \cdots \\
&\quad + Q_t R_t' L_{t+1}'\cdots L_{n-1}' Z_n' F_n^{-1} Z_n L_{n-1}\cdots L_{j+1} \\
&= Q_t R_t' L_{t+1}'\cdots L_{j-1}' L_j' N_j, \qquad (4.77)
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}(\alpha_t r_j') &= \mathrm{E}(\alpha_t v_{j+1}')F_{j+1}^{-1}Z_{j+1} + \mathrm{E}(\alpha_t v_{j+2}')F_{j+2}^{-1}Z_{j+2}L_{j+1} + \cdots \\
&\quad + \mathrm{E}(\alpha_t v_n')F_n^{-1}Z_n L_{n-1}\cdots L_{j+1} \\
&= P_t L_t' L_{t+1}'\cdots L_j' Z_{j+1}' F_{j+1}^{-1} Z_{j+1} \\
&\quad + P_t L_t' L_{t+1}'\cdots L_{j+1}' Z_{j+2}' F_{j+2}^{-1} Z_{j+2} L_{j+1} + \cdots \\
&\quad + P_t L_t' L_{t+1}'\cdots L_{n-1}' Z_n' F_n^{-1} Z_n L_{n-1}\cdots L_{j+1} \\
&= P_t L_t' L_{t+1}'\cdots L_{j-1}' L_j' N_j, \qquad (4.78)
\end{aligned}$$

for $j = t, \ldots, n$. Hence

$$\begin{aligned}
\mathrm{E}(\varepsilon_t r_j') &= \mathrm{E}(\varepsilon_t x_{t+1}')N_{t+1,j}^*, \\
\mathrm{E}(\eta_t r_j') &= \mathrm{E}(\eta_t x_{t+1}')N_{t+1,j}^*, \\
\mathrm{E}(\alpha_t r_j') &= \mathrm{E}(\alpha_t x_{t+1}')N_{t+1,j}^*, \qquad (4.79)
\end{aligned}$$

where $N_{t,j}^* = L_t'\cdots L_{j-1}'L_j'N_j$ for $j = t, \ldots, n$.

The cross-expectations of ε_t, η_t and α_t between the smoothed estimators

$$\hat{\varepsilon}_j = H_j\big(F_j^{-1}v_j - K_j'r_j\big), \qquad \hat{\eta}_j = Q_j R_j' r_j, \qquad \alpha_j - \hat{\alpha}_j = x_j - P_j r_{j-1},$$

for $j = t+1, \ldots, n$, are given by

$$
\begin{aligned}
\mathrm{E}(\varepsilon_t \hat{\varepsilon}_j') &= \mathrm{E}(\varepsilon_t v_j') F_j^{-1} H_j - \mathrm{E}(\varepsilon_t r_j') K_j H_j, \\
\mathrm{E}(\varepsilon_t \hat{\eta}_j') &= \mathrm{E}(\varepsilon_t r_j') R_j Q_j, \\
\mathrm{E}[\varepsilon_t (\alpha_j - \hat{\alpha}_j)'] &= \mathrm{E}(\varepsilon_t x_j') - \mathrm{E}(\varepsilon_t r_{j-1}') P_j, \\
\mathrm{E}(\eta_t \hat{\varepsilon}_j') &= \mathrm{E}(\eta_t v_j') F_j^{-1} H_j - \mathrm{E}(\eta_t r_j') K_j H_j, \\
\mathrm{E}(\eta_t \hat{\eta}_j') &= \mathrm{E}(\eta_t r_j') R_j Q_j, \\
\mathrm{E}[\eta_t (\alpha_j - \hat{\alpha}_j)'] &= \mathrm{E}(\eta_t x_j') - \mathrm{E}(\eta_t r_{j-1}') P_j, \\
\mathrm{E}(\alpha_t \hat{\varepsilon}_j') &= \mathrm{E}(\alpha_t v_j') F_j^{-1} H_j - \mathrm{E}(\alpha_t r_j') K_j H_j, \\
\mathrm{E}(\alpha_t \hat{\eta}_j') &= \mathrm{E}(\alpha_t r_j') R_j Q_j, \\
\mathrm{E}[\alpha_t (\alpha_j - \hat{\alpha}_j)'] &= \mathrm{E}(\alpha_t x_j') - \mathrm{E}(\alpha_t r_{j-1}') P_j,
\end{aligned}
$$

into which the expressions in equations (4.75), (4.76), (4.77) and (4.78) can be substituted.

The covariance matrices of the smoothed estimators at different times are derived as follows. We first consider the covariance matrix for the smoothed disturbance vector $\hat{\varepsilon}_t$, that is, $\mathrm{Cov}(\varepsilon_t - \hat{\varepsilon}_t, \varepsilon_j - \hat{\varepsilon}_j)$ for $t = 1, \ldots, n$ and $j = t+1, \ldots, n$. Since

$$
\mathrm{E}[\hat{\varepsilon}_t (\varepsilon_j - \hat{\varepsilon}_j)'] = \mathrm{E}[\mathrm{E}\{\hat{\varepsilon}_t (\varepsilon_j - \hat{\varepsilon}_j)' | Y_n\}] = 0,
$$

we have

$$
\begin{aligned}
\mathrm{Cov}(\varepsilon_t - \hat{\varepsilon}_t, \varepsilon_j - \hat{\varepsilon}_j) &= \mathrm{E}[\varepsilon_t (\varepsilon_j - \hat{\varepsilon}_j)'] \\
&= -\mathrm{E}(\varepsilon_t \hat{\varepsilon}_j') \\
&= H_t K_t' L_{t+1}' \cdots L_{j-1}' Z_j' F_j^{-1} H_j \\
&\quad + H_t K_t' L_{t+1}' \cdots L_{j-1}' L_j' N_j K_j H_j \\
&= H_t K_t' L_{t+1}' \cdots L_{j-1}' W_j',
\end{aligned}
$$

where

$$
W_j = H_j \big(F_j^{-1} Z_j - K_j' N_j L_j \big), \tag{4.80}
$$

for $j = t+1, \ldots, n$. In a similar way obtain

$$
\begin{aligned}
\mathrm{Cov}(\eta_t - \hat{\eta}_t, \eta_j - \hat{\eta}_j) &= -\mathrm{E}(\eta_t \hat{\eta}_j') \\
&= -Q_t R_t' L_{t+1}' \cdots L_{j-1}' L_j' N_j R_j Q_j,
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Cov}(\alpha_t - \hat{\alpha}_t, \alpha_j - \hat{\alpha}_j) &= -\mathrm{E}[\alpha_t(\alpha_j - \hat{\alpha}_j)'] \\
&= P_t L_t' L_{t+1}' \cdots L_{j-1}' - P_t L_t' L_{t+1}' \cdots L_{j-1}' N_{j-1} P_j \\
&= P_t L_t' L_{t+1}' \cdots L_{j-1}'(I - N_{j-1} P_j),
\end{aligned}
$$

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

The cross-covariance matrices of the smoothed disturbances are obtained as follows. We have

$$
\begin{aligned}
\mathrm{Cov}(\varepsilon_t - \hat{\varepsilon}_t, \eta_j - \hat{\eta}_j) &= \mathrm{E}[(\varepsilon_t - \hat{\varepsilon}_t)(\eta_j - \hat{\eta}_j)'] \\
&= \mathrm{E}[\varepsilon_t(\eta_j - \hat{\eta}_j)'] \\
&= -\mathrm{E}(\varepsilon_t \hat{\eta}_j') \\
&= H_t K_t' L_{t+1}' \cdots L_{j-1}' L_j' N_j R_j Q_j,
\end{aligned}
$$

for $j = t, t+1, \ldots, n$, and

$$
\begin{aligned}
\mathrm{Cov}(\eta_t - \hat{\eta}_t, \varepsilon_j - \hat{\varepsilon}_j) &= -\mathrm{E}(\eta_t \hat{\varepsilon}_j') \\
&= -Q_t R_t' L_{t+1}' \cdots L_{j-1}' Z_j' F_j^{-1} H_j \\
&\quad + Q_t R_t' L_{t+1}' \cdots L_{j-1}' N_j' K_j H_j \\
&= -Q_t R_t' L_{t+1}' \cdots L_{j-1}' W_j',
\end{aligned}
$$

for $j = t+1, \ldots, n$. The matrix products $L_{t+1}' \ldots L_{j-1}' L_j'$ for $j = t$ and $L_{t+1}' \ldots L_{j-1}'$ for $j = t+1$ are assumed to be equal to the identity matrix.

The cross-covariances between the smoothed state vector and the smoothed disturbances are obtained in a similar way. We have

$$
\begin{aligned}
\mathrm{Cov}(\alpha_t - \hat{\alpha}_t, \varepsilon_j - \hat{\varepsilon}_j) &= -\mathrm{E}(\alpha_t \hat{\varepsilon}_j') \\
&= -P_t L_t' L_{t+1}' \cdots L_{j-1}' Z_j' F_j^{-1} H_j \\
&\quad + P_t L_t' L_{t+1}' \cdots L_{j-1}' N_j' K_j H_j \\
&= -P_t L_t' L_{t+1}' \cdots L_{j-1}' W_j', \\
\mathrm{Cov}(\alpha_t - \hat{\alpha}_t, \eta_j - \hat{\eta}_j) &= -\mathrm{E}(\alpha_t \hat{\eta}_j') \\
&= -P_t L_t' L_{t+1}' \cdots L_{j-1}' L_j' N_j R_j Q_j,
\end{aligned}
$$

for $j = t, t+1, \ldots, n$, and

$$
\begin{aligned}
\mathrm{Cov}(\varepsilon_t - \hat{\varepsilon}_t, \alpha_j - \hat{\alpha}_j) &= \mathrm{E}[\varepsilon_t(\alpha_j - \hat{\alpha}_j)'] \\
&= -H_t K_t' L_{t+1}' \cdots L_{j-1}' \\
&\quad + H_t K_t' L_{t+1}' \cdots L_{j-1}' N_{j-1} P_j \\
&= -H_t K_t' L_{t+1}' \cdots L_{j-1}'(I - N_{j-1} P_j),
\end{aligned}
$$

Table 4.4 Covariances of smoothed estimators for $t = 1, \ldots, n$.

$\hat{\varepsilon}_t$	$\hat{\varepsilon}_j$	$H_t K_t' L_{t+1}' \cdots L_{j-1}' W_j'$	$j > t$
	$\hat{\eta}_j$	$H_t K_t' L_{t+1}' \cdots L_{j-1}' L_j' N_j R_j Q_j$	$j \geq t$
	$\hat{\alpha}_j$	$-H_t K_t' L_{t+1}' \cdots L_{j-1}'(I_m - N_{j-1} P_j)$	$j > t$
$\hat{\eta}_t$	$\hat{\varepsilon}_j$	$-Q_t R_t' L_{t+1}' \cdots L_{j-1}' W_j'$	$j > t$
	$\hat{\eta}_j$	$-Q_t R_t' L_{t+1}' \cdots L_{j-1}' L_j' N_j R_j Q_j$	$j > t$
	$\hat{\alpha}_j$	$Q_t R_t' L_{t+1}' \cdots L_{j-1}'(I_m - N_{j-1} P_j)$	$j > t$
$\hat{\alpha}_t$	$\hat{\varepsilon}_j$	$-P_t L_t' L_{t+1}' \cdots L_{j-1}' W_j'$	$j \geq t$
	$\hat{\eta}_j$	$-P_t L_t' L_{t+1}' \cdots L_{j-1}' L_j' N_j R_j Q_j$	$j \geq t$
	$\hat{\alpha}_j$	$P_t L_t' L_{t+1}' \cdots L_{j-1}'(I_m - N_{j-1} P_j)$	$j \geq t$

$$
\begin{aligned}
\mathrm{Cov}(\eta_t - \hat{\eta}_t, \alpha_j - \hat{\alpha}_j) &= \mathrm{E}[\eta_t(\alpha_j - \hat{\alpha}_j)'] \\
&= Q_t R_t' L_{t+1}' \cdots L_{j-1}'(I - N_{j-1} P_j),
\end{aligned}
$$

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

The results here have been developed by de Jong and MacKinnon (1988), who derived the covariances between smoothed state vector estimators, and by Koopman (1993) who derived the covariances between the smoothed disturbance vectors estimators. The results in this section have also been reviewed by de Jong (1998). The auto- and cross-covariance matrices are for convenience collected in Table 4.4.

4.8 Weight functions

4.8.1 Introduction

Up to this point we have developed recursions for the evaluation of the conditional mean vector and variance matrix of the state vector α_t given the observations $y_1, \ldots, y_{t-1}$ (prediction), given the observations $y_1, \ldots, y_t$ (filtering) and given the observations $y_1, \ldots, y_n$ (smoothing). We have also developed recursions for the conditional mean vectors and variance matrices of the disturbance vectors ε_t and η_t given the observation $y_1, \ldots, y_n$. It follows that these conditional means are weighted sums of past (filtering), of past and present (contemporaneous filtering) and of all (smoothing) observations. It is of interest to study these weights to gain a better understanding of the properties of the estimators as is argued in Koopman and Harvey (2003). For example, the weights for the smoothed estimator of a trend component around time $t = n/2$, that is, in the middle of the series, should be symmetric and centred around t with exponentially declining weights unless specific circumstances require a different pattern. Models which produce weight patterns for the trend components which differ from what is regarded as appropriate should be investigated. In effect, the weights can be regarded as what are known as kernel functions in the field of nonparametric regression; see, for example, Green and Silverman (1994).

In the case when the state vector contains regression coefficients, the associated weights for the smoothed state vector can be interpreted as leverage statistics as studied in Cook and Weisberg (1982) and Atkinson (1985) in the context of regression models. Such statistics for state space models have been developed with the emphasis on the smoothed signal estimator $Z_t\hat{\alpha}_t$ by, for example, Kohn and Ansley (1989), de Jong (1989), Harrison and West (1991) and de Jong (1998). Since the concept of leverage is more useful in a regression context, we will refer to the expressions below as weights. Given the results of this chapter so far, it is straightforward to develop the weight expressions.

4.8.2 Filtering weights

It follows from the linear properties of the normal distribution that the filtered estimator of the state vector can be expressed as a weighted vector sum of past observations, that is

$$a_t = \sum_{j=1}^{t-1} \omega_{jt} y_j,$$

where ω_{jt} is an $m \times p$ matrix of weights associated with the estimator a_t and the jth observation. An expression for the weight matrix can be obtained using the fact that

$$\mathrm{E}(a_t \varepsilon_j') = \omega_{jt}\mathrm{E}(y_j \varepsilon_j') = \omega_{jt} H_j.$$

Since $x_t = \alpha_t - a_t$ and $\mathrm{E}(\alpha_t \varepsilon_j') = 0$, we can use (4.29) to obtain

$$\begin{aligned}\mathrm{E}(a_t \varepsilon_j') &= \mathrm{E}(x_t \varepsilon_j') = L_{t-1}\,\mathrm{E}(x_{t-1}\varepsilon_j') \\ &= L_{t-1}L_{t-2}\cdots L_{j+1}K_j H_j,\end{aligned}$$

which gives

$$\omega_{jt} = L_{t-1}L_{t-2}\cdots L_{j+1}K_j,$$

for $j = t-1, \ldots, 1$. In a similar way we can obtain the weights associated with other filtering estimators. In Table 4.5 we give a selection of such expressions from which the weights are obtained by disregarding the last matrix H_j. Finally, the expression for weights of $Z_t a_{t|t}$ follows since $Z_t P_t Z_t' = F_t - H_t$ and

$$Z_t\left(I - P_t Z_t' F_t^{-1} Z_t\right) = \left[I - (F_t - H_t)F_t^{-1}\right] Z_t = H_t F_t^{-1} Z_t.$$

Table 4.5 Expressions for $\mathrm{E}(s_t \varepsilon_j')$ with $1 \le t \le n$ given (filtering).

s_t	$j < t$	$j = t$	$j > t$
a_t	$L_{t-1}\cdots L_{j+1}K_j H_j$	0	0
$a_{t\|t}$	$(I - P_t Z_t' F_t^{-1} Z_t)L_{t-1}\cdots L_{j+1}K_j H_j$	$P_t Z_t' F_t^{-1} H_t$	0
$Z_t a_t$	$Z_t L_{t-1}\cdots L_{j+1}K_j H_j$	0	0
$Z_t a_{t\|t}$	$H_t F_t^{-1} Z_t L_{t-1}\cdots L_{j+1}K_j H_j$	$\left(I - H_t F_t^{-1}\right) H_t$	0
v_t	$-Z_t L_{t-1}\cdots L_{j+1}K_j H_j$	H_t	0

Table 4.6 Expressions for $\mathrm{E}(s_t \varepsilon_j')$ with $1 \leq t \leq n$ given (smoothing).

s_t	$j < t$	$j = t$	$j > t$
$\hat{\varepsilon}_t$	$-W_t L_{t-1} \cdots L_{j+1} K_j H_j$	$H_t D_t H_t$	$-H_t K_t' L_{t+1}' \cdots L_{j-1}' W_j'$
$\hat{\eta}_t$	$-Q_t R_t' N_t L_t L_{t-1} \cdots L_{j+1} K_j H_j$	$Q_t R_t' N_t K_t H_t$	$Q_t R_t' L_{t+1}' \cdots L_{j-1}' W_j'$
$\hat{\alpha}_t$	$(I - P_t N_{t-1}) L_{t-1} \cdots L_{j+1} K_j H_j$	$P_t W_t'$	$P_t L_t' L_{t+1}' \cdots L_{j-1}' W_j'$
$Z_t \hat{\alpha}_t$	$W_t L_{t-1} \cdots L_{j+1} K_j H_j$	$(I - H_t D_t) H_t$	$H_t K_t' L_{t+1}' \cdots L_{j-1}' W_j'$

4.8.3 Smoothing weights

The weighting expressions for smoothing estimators can be obtained in a similar way to those used for filtering. For example, the smoothed estimator of the measurement disturbance vector can be expressed as a weighted vector sum of past, current and future observations, that is

$$\hat{\varepsilon}_t = \sum_{j=1}^{n} \omega_{jt}^{\varepsilon} y_j,$$

where $\omega_{jt}^{\varepsilon}$ is a $p \times p$ matrix of weights associated with the estimator $\hat{\varepsilon}_t$ and the jth observation. An expression for the weight matrix can be obtained using the fact that

$$\mathrm{E}(\hat{\varepsilon}_t \varepsilon_j') = \omega_{jt}^{\varepsilon}\, \mathrm{E}(y_j \varepsilon_j') = \omega_{jt}^{\varepsilon} H_j.$$

Expressions for the covariance matrices for smoothed disturbances are developed in Section 4.7 and they are directly related to the expression for $\mathrm{E}(\hat{\varepsilon}_t \varepsilon_j')$ because

$$\mathrm{Cov}(\varepsilon_t - \hat{\varepsilon}_t, \varepsilon_j - \hat{\varepsilon}_j) = \mathrm{E}[(\varepsilon_t - \hat{\varepsilon}_t)\varepsilon_j'] = -\mathrm{E}(\hat{\varepsilon}_t \varepsilon_j'),$$

with $1 \leq t \leq n$ and $j = 1, \ldots, n$. Therefore, no new derivations need to be given here and we only state the results as presented in Table 4.6.

For example, to obtain the weights for the smoothed estimator of α_t, we require

$$\begin{aligned} \mathrm{E}(\hat{\alpha}_t \varepsilon_j') &= -\mathrm{E}[(\alpha_t - \hat{\alpha}_t)\varepsilon_j'] = -\mathrm{E}[\varepsilon_j(\alpha_t - \hat{\alpha}_t)']' \\ &= \mathrm{Cov}(\varepsilon_j - \hat{\varepsilon}_j, \alpha_t - \hat{\alpha}_t)', \end{aligned}$$

for $j < t$. An expression for this latter quantity can be directly obtained from Table 4.4 but notice that the indices j and t of Table 4.4 need to be reversed here. Further,

$$\mathrm{E}(\hat{\alpha}_t \varepsilon_j') = \mathrm{Cov}(\alpha_t - \hat{\alpha}_t, \varepsilon_j - \hat{\varepsilon}_j)$$

for $j \geq t$ can also be obtained from Table 4.4. In the same way we can obtain the weights for the smoothed estimators of ε_t and η_t from Table 4.4 as reported in Table 4.6. Finally, the expression for weights of $Z_t \hat{\alpha}_t$ follows since

$Z_tP_tZ_t' = F_t - H_t$ and $Z_tP_tL_t' = H_tK_t'$. Hence, by using the equations (4.42), (4.66) and (4.80), we have

$$
\begin{aligned}
Z_t(I - P_tN_{t-1}) &= Z_t - Z_tP_tZ_t'F_t^{-1}Z_t + Z_tP_tL_t'N_tL_t \\
&= H_tF_t^{-1}Z_t + H_tK_t'N_tL_t = W_t, \\
Z_tP_tW_t' &= (Z_tP_tZ_t'F_t^{-1} - Z_tP_tL_t'N_tK_t)H_t \\
&= [(F_t - H_t)F_t^{-1} - H_tK_t'N_tK_t]H_t \\
&= (I - H_tD_t)H_t.
\end{aligned}
$$

4.9 Simulation smoothing

The drawing of samples of state or disturbance vectors conditional on the observations held fixed is called *simulation smoothing*. Such samples are useful for investigating the performance of techniques of analysis proposed for the linear Gaussian model, and for Bayesian analysis based on this model. The primary purpose of simulation smoothing in this book, however, will be to serve as the basis for the simulation techniques we shall develop in Part II for dealing with non-Gaussian and nonlinear models from both classical and Bayesian perspectives.

In this section we will show how to draw random samples of the disturbance vectors ε_t and η_t, and the state vector α_t, for $t = 1, \dots, n$, generated by the linear Gaussian model (4.12) conditional on the observed vector y_n. The resulting algorithm is sometimes called a *forwards filtering, backwards sampling* algorithm.

Frühwirth-Schnatter (1994) and Carter and Kohn (1994) independently developed methods for simulation smoothing of the state vector based on the identity

$$
p(\alpha_1, \dots, \alpha_n|Y_n) = p(\alpha_n|Y_n)p(\alpha_{n-1}|Y_n, \alpha_n) \cdots p(\alpha_1|Y_n, \alpha_2, \dots, \alpha_n). \quad (4.81)
$$

de Jong and Shephard (1995) made significant progress by first concentrating on sampling the disturbances and subsequently sampling the states.

Subsequently, in Durbin and Koopman (2002), we developed a method which is based only on mean corrections of unconditional vectors and which is much simpler and computationally more efficient than the de Jong-Shephard and earlier procedures. The treatment which follows is based on this approach; the de Jong-Shephard method is summarised in Subsection 4.9.3.

4.9.1 Simulation smoothing by mean corrections

Our aim is to draw samples of the disturbances $\varepsilon_1, \dots, \varepsilon_n$ and $\eta_1, \dots, \eta_n$ given the observational set Y_n. Let $w = (\varepsilon_1', \eta_1', \dots, \varepsilon_n', \eta_n')'$ and let $\widehat{w} = \mathrm{E}(w|Y_n)$, $W = \mathrm{Var}(w|Y_n)$. Since model (4.12) is linear and Gaussian, the conditional density of w given Y_n is $\mathrm{N}(\widehat{w}, W)$. The mean vector $\widehat{w}$ is easily calculated from recursions

(4.58) and (4.63); we show below that for the mean-correction method we do not need to calculate the variance matrix W, which is convenient computationally.

The unconditional distribution of w is

$$p(w) = \mathrm{N}(0, \Phi), \qquad \text{where} \quad \Phi = \mathrm{diag}(H_1, Q_1, \ldots, H_n, Q_n).$$

Let w^+ be a random vector drawn from $p(w)$. The process of drawing w^+ is straightforward, particularly since in many cases in practice the matrices H_t and Q_t for $t = 1, \ldots, n$ are scalar or diagonal. Denote by y^+ the stacked vector of values of y_t generated recursively by drawing a vector α_1^+ from $p(\alpha_1)$, assuming this density is known, and replacing α_1 and w in model (4.12) by α_1^+ and w^+. Compute $\widehat{w}^+ = \mathrm{E}(w|y^+)$ from recursions (4.58) and (4.63). It follows from Lemma 1 of Section 4.2 that the conditional variance matrix of a vector x given another vector y in a multivariate normal distribution does not depend on the value of y. Hence, $\mathrm{Var}(w|y^+) = W$ and conditionally on y^+ we have $w^+ - \widehat{w}^+ \sim \mathrm{N}(0, W)$. Since the density $\mathrm{N}(0, W)$ does not depend on y^+, it holds that $w^+ - \widehat{w}^+ \sim \mathrm{N}(0, W)$ unconditionally, and

$$\widetilde{w} = w^+ - \widehat{w}^+ + \widehat{w}, \tag{4.82}$$

is a random draw from $\mathrm{N}(\widehat{w}, W)$ as required. The simplicity of this expression for $\widetilde{w}$, and the elementary nature of the draw of w^+ from $\mathrm{N}(0, \Phi)$, together account for the greater efficiency of this approach to simulation smoothing compared with earlier methods. Of course, if draws of $\varepsilon = (\varepsilon_1', \ldots, \varepsilon_n')'$ only or $\eta = (\eta_1', \ldots, \eta_n')'$ only are required we just replace w by ε or η as appropriate. It is interesting to note that the form of (4.82) is similar to expression (4) of Journel (1974); however, Journel's work was done in a different context. Also, we were unaware of it when developing the mean-corrections method in Durbin and Koopman (2002).

The above theory has been derived on the assumption that the initial vector α_1 has the distribution $\alpha_1 \sim \mathrm{N}(a_1, P_1)$ where a_1 and P_1 are known. In practice, however, it is common for at least some of the elements of α_1 to be either fixed and unknown or to be random variables with arbitrarily large variances; such elements are called *diffuse*. The modifications to the theory that are required when some elements of α_1 are diffuse are given in Section 5.5.

4.9.2 Simulation smoothing for the state vector

To construct an algorithm for generating draws of the state vector $\alpha = (\alpha_1', \ldots, \alpha_n')'$ from the conditional density $p(\alpha|Y_n)$, we denote a draw from $p(\alpha)$ as α^+ and a draw from $p(\alpha|Y_n)$ as $\widetilde{\alpha}$. To generate α^+ we first draw w^+ as above and then use model (4.12) as a recursion initialised by $\alpha_1^+ \sim p(\alpha_1)$ with α and w replaced by α^+ and w^+, respectively, as in Subsection 4.9.1. We compute $\widehat{\alpha} = \mathrm{E}(\alpha|Y_n)$ and $\widehat{\alpha}^+ = \mathrm{E}(\alpha|y^+)$ by the Kalman filter and the smoothing recursions (4.58) and (4.63), finally using the forward recursion (4.71) to generate $\widehat{\alpha}$ and $\widehat{\alpha}^+$. The required draw of $\widetilde{\alpha}$ is then given by the expression $\widetilde{\alpha} = \alpha^+ - \widehat{\alpha}^+ + \widehat{\alpha}$.

4.9.3 de Jong–Shephard method for simulation of disturbances

The mean-corrections method for simulation smoothing works well in nearly all practical cases. However, there may be cases where the mean-corrections method cannot be implemented properly due to imposed ill-defined variance matrices; see the discussion in Jungbacker and Koopman (2007, §1). Since the de Jong–Shephard method does work generally, we present here the recursions developed in de Jong and Shephard (1995). We begin by presenting the recursions required for drawing a sample of the observation disturbances $\varepsilon_1, \dots, \varepsilon_n$ from the conditional density $p(\varepsilon_1, \dots, \varepsilon_n | Y_n)$. Let

$$\bar{\varepsilon}_t = \mathrm{E}(\varepsilon_t | \varepsilon_{t+1}, \dots, \varepsilon_n, Y_n), \qquad t = n-1, \dots, 1, \tag{4.83}$$

with $\bar{\varepsilon}_n = \mathrm{E}(\varepsilon_n | Y_n) = H_n F_n^{-1} v_n$. It can be shown that

$$\bar{\varepsilon}_t = H_t \left(F_t^{-1} v_t - K_t' \tilde{r}_t \right), \qquad t = n-1, \dots, 1, \tag{4.84}$$

where $\tilde{r}_t$ is determined by the backward recursion

$$\tilde{r}_{t-1} = Z_t' F_t^{-1} v_t - \tilde{W}_t' C_t^{-1} d_t + L_t' \tilde{r}_t, \qquad t = n, n-1, \dots, 1, \tag{4.85}$$

with $\tilde{r}_n = 0$ and

$$\tilde{W}_t = H_t \left(F_t^{-1} Z_t - K_t' \tilde{N}_t L_t \right), \tag{4.86}$$

$$\tilde{N}_{t-1} = Z_t' F_t^{-1} Z_t + \tilde{W}_t' C_t^{-1} \tilde{W}_t + L_t' \tilde{N}_t L_t, \tag{4.87}$$

for $t = n, n-1, \dots, 1$ and with $\tilde{N}_n = 0$. Here $C_t = \mathrm{Var}(\varepsilon_t - \bar{\varepsilon}_t)$ which is determined by

$$C_t = H_t - H_t \left(F_t^{-1} + K_t' \tilde{N}_t K_t \right) H_t, \qquad t = n, \dots, 1. \tag{4.88}$$

In these formulae, F_t, v_t and K_t are obtained from the Kalman filter (4.24) and $L_t = T_t - K_t Z_t$. The required draw from $p(\varepsilon_t | \varepsilon_{t+1}, \dots, \varepsilon_n, Y_n)$ is then obtained as a random draw from $\mathrm{N}(\bar{\varepsilon}_t, C_t)$.

We now present the recursions required for selecting a sample of state disturbances $\eta_1, \dots, \eta_n$ from density $p(\eta_1, \dots, \eta_n | Y_n)$. Let

$$\bar{\eta}_t = \mathrm{E}(\eta_t | \eta_{t+1}, \dots, \eta_n, Y_n), \qquad \bar{C}_t = \mathrm{Var}(\eta_t | Y_n, \eta_{t+1}, \dots, \eta_n),$$

for $t = n-1, \dots, 1$ with $\bar{\eta}_n = \mathrm{E}(\eta_n | Y_n) = 0$ and $\bar{C}_n = \mathrm{Var}(\eta_n | Y_n) = Q_n$. Further, let

$$\bar{W}_t = Q_t R_t' \tilde{N}_t L_t, \tag{4.89}$$

where $\tilde{N}_t$ is determined by the backward recursion (4.87) with $\tilde{W}_t$ replaced by $\bar{W}_t$ in (4.89). Then $\bar{\eta}_t$ is given by the relation

$$\bar{\eta}_t = Q_t R_t \tilde{r}_t, \qquad t = n-1, \dots, 1, \tag{4.90}$$

where $\tilde{r}_t$ is determined by the recursion (4.85) with $\tilde{W}_t$ replaced by $\bar{W}_t$. Furthermore, $\bar{C}_t$ is given by

$$\bar{C}_t = Q_t - Q_t R_t' \tilde{N}_t R_t Q_t, \qquad t = n, \dots, 1. \tag{4.91}$$

with $\tilde{N}_t$ as in (4.87). The required draw from $p(\eta_t | \eta_{t+1}, \dots, \eta_n, Y_n)$ is then obtained as a random draw from $\mathrm{N}(\bar{\eta}_t, \bar{C}_t)$ for $t = n-1, \dots, 1$ with $\eta_n \sim \mathrm{N}(0, Q_n)$.

By adopting the same arguments as we used for developing the quick state smoother in Subsection 4.6.2, we obtain the following forwards recursion for simulating from the conditional density $p(\alpha | Y_n)$ when the de Jong–Shephard method is used

$$\tilde{\alpha}_{t+1} = T_t \tilde{\alpha}_t + R_t \tilde{\eta}_t, \tag{4.92}$$

for $t = 1, \dots, n$ with $\tilde{\alpha}_1 = a_1 + P_1 \tilde{r}_0$.

The proofs of these recursions are long and intricate so they will not be given here. Instead, the reader is referred to the proofs in de Jong and Shephard (1995, §2) and Jungbacker and Koopman (2007, Theorem 2 and Proposition 6).

4.10 Missing observations

We now demonstrate that when the linear state space model (4.12) is used for the analysis, with or without the assumption of normality, allowance for missing observations in the derivation of the Kalman filter and smoother is particularly simple. Suppose that observations y_j are missing for $j = \tau, \dots, \tau^*$ with $1 < \tau < \tau^* < n$. An obvious procedure is to define a new series $y_{t^*}^*$ where $y_t^* = y_t$ for $t = t^* = 1, \dots, \tau - 1$ and $y_{t^*}^* = y_t$ for $t = \tau^* + 1, \dots, n$ and $t^* = \tau, \dots, n - (\tau^* - \tau)$. The model for $y_{t^*}^*$ is then the same as (4.12) where $y_t = y_{t^*}$, $\alpha_t = \alpha_{t^*}$ and the disturbances are associated with time index t^*. The system matrices remain associated with the time index t. The state update equation at time $t^* = \tau - 1$ is replaced by

$$\alpha_\tau = T_{\tau^*, \tau-1}^* \alpha_{\tau-1} + \eta_{\tau-1}^*, \qquad \eta_{\tau-1}^* \sim N\left(0, \sum_{j=\tau}^{\tau^*+1} T_{\tau^*, j}^* R_{j-1} Q_{j-1} R_{j-1}' T_{\tau^*, j}^{*\prime}\right),$$

with $T_{i,j}^* = T_i T_{i-1} \dots T_j$ for $j = \tau, \dots, \tau^*$ and $T_{\tau^*, \tau^*+1}^* = I_r$. Filtering and smoothing then proceed by the methods developed above for model (4.12). The procedure is extended in an obvious way when observations are missing at several points in the series.

It is, however, easier to proceed as follows. For $t = \tau, \dots, \tau^* - 1$ we have

$$\begin{aligned} a_{t|t} &= \mathrm{E}(\alpha_t|Y_t) = \mathrm{E}(\alpha_t|Y_{t-1}) \\ &= a_t, \\ P_{t|t} &= \mathrm{Var}(\alpha_t|Y_t) = \mathrm{Var}(\alpha_t|Y_{t-1}) \\ &= P_t, \\ a_{t+1} &= \mathrm{E}(\alpha_{t+1}|Y_t) = \mathrm{E}(T_t\alpha_t + R_t\eta_t|Y_{t-1}) \\ &= T_t a_t, \\ P_{t+1} &= \mathrm{Var}(\alpha_{t+1}|Y_t) = \mathrm{Var}(T_t\alpha_t + R_t\eta_t|Y_{t-1}) \\ &= T_t P_t T_t' + R_t Q_t R_t'. \end{aligned}$$

It follows that the Kalman filter for the case of missing observations is obtained simply by putting $Z_t = 0$ in (4.24) for $t = \tau, \dots, \tau^* - 1$; the same applies to (4.26) for the case of models with mean adjustments. Similarly, the backwards smoothing recursions (4.38) and (4.42) become

$$r_{t-1} = T_t' r_t, \qquad N_{t-1} = T_t' N_t T_t, \qquad t = \tau^* - 1, \dots, \tau; \tag{4.93}$$

other relevant equations in (4.44) remain the same. It therefore follows that in smoothing as in filtering, we can use the same recursion (4.44) as when all observations are available by taking $Z_t = 0$ at time points where observations are missing. As with Kalman filtering and smoothing for complete sets of observations, the results remain valid for MLVUE and Bayesian analysis by Lemmas 2, 3 and 4. This simple treatment of missing observations is one of the attractions of the state space methods for time series analysis.

Suppose that at time t some but not all of the elements of the observation vector y_t are missing. Let y_t^* be the vector of values actually observed. Then $y_t^* = W_t y_t$ where W_t is a known matrix whose rows are a subset of the rows of I. Consequently, at time points where not all elements of y_t are available, the first equation of (4.12) is replaced by the equation

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

where $Z_t^* = W_t Z_t$, $\varepsilon_t^* = W_t \varepsilon_t$ and $H_t^* = W_t H_t W_t'$. The Kalman filter and smoother then proceed exactly as in the standard case, provided that y_t, Z_t and H_t are replaced by y_t^*, Z_t^* and H_t^* at relevant time points. Of course, the dimensionality of the observation vector varies over time, but this does not affect the validity of the formulae; see also Section 4.12. The missing elements can be estimated by appropriate elements of $Z_t\hat{\alpha}_t$ where $\hat{\alpha}_t$ is the smoothed value. A more convenient method for dealing with missing elements for such multivariate models is given in Section 6.4 which is based on an element by element treatment of the observation vector y_t.

When observations or observational elements are missing, simulation samples obtained by the methods described in Section 4.9 carry through without further complexities. The mean correction methods are based on Kalman filtering and smoothing methods which can deal with missing values in the way shown in this subsection.

4.11 Forecasting

For many time series investigations, forecasting of future observations of the state vector is of special importance. In this section, we shall show that minimum mean square error forecasts can be obtained simply by treating future values of y_t as missing observations and using the techniques of the last section.

Suppose we have vector observations $y_1, \dots, y_n$ which follow the state space model (4.12) and we wish to forecast y_{n+j} for $j = 1, \dots, J$. For this purpose let us choose the estimate $\bar{y}_{n+j}$ which has a minimum mean square error matrix given Y_n, that is, $\bar{F}_{n+j} = \mathrm{E}[(\bar{y}_{n+j} - y_{n+j})(\bar{y}_{n+j} - y_{n+j})'|Y_n]$ is a minimum in the matrix sense for all estimates of y_{n+j}. It is standard knowledge that if x is a random vector with mean μ and finite variance matrix, then the value of a constant vector λ which minimises $\mathrm{E}[(\lambda - x)(\lambda - x)']$ is $\lambda = \mu$. It follows that the minimum mean square error forecast of y_{n+j} given Y_n is the conditional mean $\bar{y}_{n+j} = \mathrm{E}(Y_{n+j}|Y_n)$.

For $j = 1$ the forecast is straightforward. We have $y_{n+1} = Z_{n+1}\alpha_{n+1} + \varepsilon_{n+1}$ so

$$
\begin{aligned}
\bar{y}_{n+1} &= Z_{n+1}\,\mathrm{E}(\alpha_{n+1}|Y_n) \\
&= Z_{n+1}a_{n+1},
\end{aligned}
$$

where a_{n+1} is the estimate (4.21) of α_{n+1} produced by the Kalman filter. The conditional mean square error matrix

$$
\begin{aligned}
\bar{F}_{n+1} &= \mathrm{E}[(\bar{y}_{n+1} - y_{n+1})(\bar{y}_{n+1} - y_{n+1})'|Y_n] \\
&= Z_{n+1}P_{n+1}Z_{n+1}' + H_{n+1},
\end{aligned}
$$

is produced by the Kalman filter relation (4.16). We now demonstrate that we can generate the forecasts $\bar{y}_{n+j}$ for $j = 2, \dots, J$ merely by treating $y_{n+1}, \dots, y_{n+J}$ as missing values as in Section 4.10. Let $\bar{a}_{n+j} = \mathrm{E}(\alpha_{n+j}|Y_n)$ and $\bar{P}_{n+j} = \mathrm{E}[(\bar{a}_{n+j} - \alpha_{n+j})(\bar{a}_{n+j} - \alpha_{n+j})'|Y_n]$. Since $y_{n+j} = Z_{n+j}\alpha_{n+j} + \varepsilon_{n+j}$ we have

$$
\begin{aligned}
\bar{y}_{n+j} &= Z_{n+j}\,\mathrm{E}(\alpha_{n+j}|Y_n) \\
&= Z_{n+j}\bar{a}_{n+j},
\end{aligned}
$$

with conditional mean square error matrix

$$
\begin{aligned}
\bar{F}_{n+j} &= \mathrm{E}[\{Z_{n+j}(\bar{a}_{n+j} - \alpha_{n+j}) - \varepsilon_{n+j}\}\{Z_{n+j}(\bar{a}_{n+j} - \alpha_{n+j}) - \varepsilon_{n+j}\}'|Y_n] \\
&= Z_{n+j}\bar{P}_{n+j}Z_{n+j}' + H_{n+j}.
\end{aligned}
$$

We now derive recursions for calculating $\bar{a}_{n+j}$ and $\bar{P}_{n+j}$. We have $\alpha_{n+j+1} = T_{n+j}\alpha_{n+j} + R_{n+j}\eta_{n+j}$ so

$$\begin{aligned}\bar{a}_{n+j+1} &= T_{n+j}\,\mathrm{E}(\alpha_{n+j}|Y_n)\\ &= T_{n+j}\bar{a}_{n+j},\end{aligned}$$

for $j = 1, \ldots, J-1$ and with $\bar{a}_{n+1} = a_{n+1}$. Also,

$$\begin{aligned}\bar{P}_{n+j+1} &= \mathrm{E}[(\bar{a}_{n+j+1} - \alpha_{n+j+1})(\bar{a}_{n+j+1} - \alpha_{n+j+1})'|Y_n]\\ &= T_{n+j}\,\mathrm{E}[(\bar{a}_{n+j} - \alpha_{n+j})(\bar{a}_{n+j} - \alpha_{n+j})'|Y_n]T'_{n+j}\\ &\quad + R_{n+j}\,\mathrm{E}[\eta_{n+j}\eta'_{n+j}]R'_{n+j}\\ &= T_{n+j}\bar{P}_{n+j}T'_{n+j} + R_{n+j}Q_{n+j}R'_{n+j},\end{aligned}$$

for $j = 1, \ldots, J-1$.

We observe that the recursions for $\bar{a}_{n+j}$ and $\bar{P}_{n+j}$ are the same as the recursions for a_{n+j} and P_{n+j} of the Kalman filter (4.24) provided that we take $Z_{n+j} = 0$ for $j = 1, \ldots, J-1$. But this is precisely the condition that in Section 4.10 enabled us to deal with missing observations by routine application of the Kalman filter. We have therefore demonstrated that forecasts of $y_{n+1}, \ldots, y_{n+J}$ together with their forecast error variance matrices can be obtained merely by treating y_t for $t > n$ as missing observations and using the results of Section 4.10. In a sense this conclusion could be regarded as intuitively obvious; however, we thought it worthwhile demonstrating it algebraically. To sum up, forecasts and their associated error variance matrices can be obtained routinely in state space time series analysis based on linear Gaussian models by continuing the Kalman filter beyond $t = n$ with $Z_t = 0$ for $t > n$. Of course, for the computation of $\bar{y}_{n+j}$ and $\bar{F}_{n+j}$ we take Z_{n+j} as their actual values for $j = 1, \ldots, J$. Similar results hold for forecasting values of the state vector α_t and hence for forecasting linear functions of elements of α_t. The results remain valid for MVLUE forecasting in the non-normal case and for Bayesian analysis using Lemmas 2 to 4. These results for forecasting are a particularly elegant feature of state space methods for time series analysis.

4.12 Dimensionality of observational vector

Throughout this chapter we have assumed, both for convenience of exposition and also because this is by far the most common case in practice, that the dimensionality of the observation vector y_t is a fixed value p. It is easy to verify, however, that none of the basic formulae that we have derived depend on this assumption. For example, the filtering recursion (4.24) and the disturbance smoother (4.69) both remain valid when the dimensionality of y_t is allowed to vary. This convenient generalisation arises because of the recursive nature of the formulae. In fact we made use of this property in the treatment of missing observational elements in Section 4.10. We do not, therefore, consider explicitly in this book the situation where the dimensionality of the state vector α_t varies with t, apart from the treatment of missing observations just referred to and the conversion of multivariate series to univariate series in Section 6.4.

4.13 Matrix formulations of basic results

In this section we provide matrix expressions for the state space model, filtering, smoothing and the associated unconditional and conditional densities. These expressions can give some additional insights into the filtering and smoothing results of this chapter. We further develop these expressions for reference purposes for the remaining chapters of this book, particularly in Part II.

4.13.1 State space model in matrix form

The linear Gaussian state space model (4.12) can itself be represented in a general matrix form. The observation equation can be formulated as

$$Y_n = Z\alpha + \varepsilon, \qquad \varepsilon \sim \mathrm{N}(0, H), \tag{4.94}$$

with

$$Y_n = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \qquad Z = \begin{bmatrix} Z_1 & & 0 & 0 \\ & \ddots & & \vdots \\ 0 & & Z_n & 0 \end{bmatrix}, \qquad \alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \\ \alpha_{n+1} \end{pmatrix},$$

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}, \qquad H = \begin{bmatrix} H_1 & & 0 \\ & \ddots & \\ 0 & & H_n \end{bmatrix}. \tag{4.95}$$

The state equation takes the form

$$\alpha = T(\alpha_1^* + R\eta), \qquad \eta \sim \mathrm{N}(0, Q), \tag{4.96}$$

with

$$T = \begin{bmatrix} I & 0 & 0 & 0 & & 0 & 0 \\ T_1 & I & 0 & 0 & & 0 & 0 \\ T_2T_1 & T_2 & I & 0 & & 0 & 0 \\ T_3T_2T_1 & T_3T_2 & T_3 & I & & 0 & 0 \\ & & & & \ddots & & \vdots \\ T_{n-1}\cdots T_1 & T_{n-1}\cdots T_2 & T_{n-1}\cdots T_3 & T_{n-1}\cdots T_4 & & I & 0 \\ T_n\cdots T_1 & T_n\cdots T_2 & T_n\cdots T_3 & T_n\cdots T_4 & \cdots & T_n & I \end{bmatrix}, \tag{4.97}$$

$$\alpha_1^* = \begin{pmatrix} \alpha_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad R = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ R_1 & 0 & & 0 \\ 0 & R_2 & & 0 \\ & & \ddots & \vdots \\ 0 & 0 & \cdots & R_n \end{bmatrix}, \tag{4.98}$$

$$\eta = \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_n \end{pmatrix}, \qquad Q = \begin{bmatrix} Q_1 & & 0 \\ & \ddots & \\ 0 & & Q_n \end{bmatrix}. \tag{4.99}$$

This representation of the state space model is useful for getting a better understanding of some of the results in this chapter. For example, it follows that $\mathrm{E}(\alpha_1^*) = a_1^*$, $\mathrm{Var}(\alpha_1^*) = P_1^*$,

$$\begin{aligned} \mathrm{E}(\alpha) = Ta_1^*, \qquad \mathrm{Var}(\alpha) &= \mathrm{Var}\left\{T\left[(\alpha_1^* - a_1^*) + R\eta\right]\right\} \\ &= T(P_1^* + RQR')T' \\ &= TQ^*T', \end{aligned} \tag{4.100}$$

where

$$a_1^* = \begin{pmatrix} a_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad P_1^* = \begin{bmatrix} P_1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & & 0 \\ 0 & 0 & 0 & & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & & 0 \end{bmatrix}, \qquad Q^* = P_1^* + RQR'.$$

Furthermore, we show that the observation vectors y_t are linear functions of the initial state vector α_1 and the disturbance vectors ε_t and η_t for $t = 1, \ldots, n$ since it follows by substitution of (4.96) into (4.94) that

$$Y_n = ZT\alpha_1^* + ZTR\eta + \varepsilon. \tag{4.101}$$

It also follows that

$$\begin{aligned} \mathrm{E}(Y_n) = \mu = ZTa_1^*, \qquad \mathrm{Var}(Y_n) = \Omega &= ZT(P_1^* + RQR')T'Z' + H \\ &= ZTQ^*T'Z' + H. \end{aligned} \tag{4.102}$$

4.13.2 Matrix expression for densities

Given the model expressions in matrix form, we can also express the densities of $p(Y_n)$ and $p(\alpha, Y_n)$ in terms of vectors and matrices. For example, it follows from (4.102) that the log of the density function $p(Y_n)$ is given by

$$\log p(Y_n) = \text{constant} - \frac{1}{2}\log|\,\Omega\,| - \frac{1}{2}(Y_n - \mu)'\Omega^{-1}(Y_n - \mu). \tag{4.103}$$

An expression for the joint density of Y_n and α can be based on the decomposition $p(\alpha, Y_n) = p(Y_n|\alpha)p(\alpha)$. It follows from (4.100) that the logdensity of α is given by

$$\log p(\alpha) = \text{constant} - \frac{1}{2}\log|V^*| - \frac{1}{2}(\alpha - a^*)'V^{*\,-1}(\alpha - a^*), \tag{4.104}$$

where $a^* = \mathrm{E}(\alpha) = Ta_1^*$ and $V^* = \mathrm{Var}(\alpha) = TQ^*T'$. The observation equation (4.94) implies that the logdensity of the observation vector Y_n given the state α is given by

$$\log p(Y_n|\alpha) = \text{constant} - \frac{1}{2}\log|H| - \frac{1}{2}(Y_n - \theta)'H^{-1}(Y_n - \theta), \tag{4.105}$$

where $\theta = Z\alpha$ is referred to as the signal. It follows that

$$p(Y_n|\alpha) = p(Y_n|\theta).$$

The joint logdensity $\log p(\alpha, Y_n)$ is simply the sum of $\log p(\alpha)$ and $\log p(Y_n|\theta)$.

4.13.3 Filtering in matrix form: Cholesky decomposition

We now show that the vector of innovations can be represented as $v = CY_n - Ba_1^*$ where $v = (v_1', \ldots, v_n')'$ and where C and B are matrices of which C is lower block triangular. First we observe that

$$a_{t+1} = L_t a_t + K_t y_t,$$

which follows from (4.21) with $v_t = y_t - Z_t a_t$ and $L_t = T_t - K_t Z_t$. Then by substituting repeatedly we have

$$a_{t+1} = L_t L_{t-1} \cdots L_1 a_1 + \sum_{j=1}^{t-1} L_t L_{t-1} \cdots L_{j+1} K_j y_j + K_t y_t$$

and

$$\begin{aligned} v_1 &= y_1 - Z_1 a_1, \\ v_2 &= -Z_2 L_1 a_1 + y_2 - Z_2 K_1 y_1, \\ v_3 &= -Z_3 L_2 L_1 a_1 + y_3 - Z_3 K_2 y_2 - Z_3 L_2 K_1 y_1, \end{aligned}$$

and so on. Generally,

$$\begin{aligned} v_t = &-Z_t L_{t-1} L_{t-2} \cdots L_1 a_1 + y_t - Z_t K_{t-1} y_{t-1} \\ &- Z_t \sum_{j=1}^{t-2} L_{t-1} \cdots L_{j+1} K_j y_j. \end{aligned}$$

Note that the matrices K_t and L_t depend on P_1, Z, T, R, H and Q but do not depend on the initial mean vector a_1 or the observations $y_1, \ldots, y_n$, for $t = 1, \ldots, n$. The innovations can thus be represented as

$$\begin{aligned} v &= (I - ZLK)Y_n - ZLa_1^* \\ &= CY_n - Ba_1^*, \end{aligned} \tag{4.106}$$

where $C = I - ZLK$, $B = ZL$,

$$L = \begin{bmatrix} I & 0 & 0 & 0 & & 0 & 0 \\ L_1 & I & 0 & 0 & & 0 & 0 \\ L_2L_1 & L_2 & I & 0 & & 0 & 0 \\ L_3L_2L_1 & L_3L_2 & L_3 & I & & 0 & 0 \\ & & & & \ddots & & \vdots \\ L_{n-1}\cdots L_1 & L_{n-1}\cdots L_2 & L_{n-1}\cdots L_3 & L_{n-1}\cdots L_4 & & I & 0 \\ L_n\cdots L_1 & L_n\cdots L_2 & L_n\cdots L_3 & L_n\cdots L_4 & \cdots & L_n & I \end{bmatrix},$$

$$K = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ K_1 & 0 & \cdots & 0 \\ 0 & K_2 & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & K_n \end{bmatrix},$$

and matrix Z is defined in (4.95). It can be easily verified that matrix C is lower block triangular with identity matrices on the leading diagonal blocks. Since $v = CY_n - ZLa_1^*$, $\mathrm{Var}(Y_n) = \Omega$ and a_1^* is constant, it follows that $\mathrm{Var}(v) = C\Omega C'$. However, we know from Subsection 4.3.5 that the innovations are independent of each other so that $\mathrm{Var}(v)$ is the block diagonal matrix

$$F = \begin{bmatrix} F_1 & 0 & \cdots & 0 \\ 0 & F_2 & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & F_n \end{bmatrix}.$$

This shows that in effect the Kalman filter is essentially equivalent to a block version of a Cholesky decomposition applied to the observational variance matrix implied by the state space model (4.12). A general discussion of the Cholesky decomposition is provided by Golub and Van Loan (1996, §4.2).

Given the special structure of $C = I - ZLK$ we can reproduce some interesting results in matrix form. We first notice from (4.106) that $\mathrm{E}(v) = C\mathrm{E}(Y_n) - ZLa_1^* = 0$. Since $\mathrm{E}(Y_n) = \mu = ZTa_1^*$ from (4.102), we obtain the identity $CZT = ZL$. It follows that

$$v = C(Y_n - \mu), \qquad F = \mathrm{Var}(v) = C\Omega C'. \tag{4.107}$$

Further we notice that C is nonsingular and

$$\Omega^{-1} = C'F^{-1}C. \tag{4.108}$$

It can be verified that matrix $C = I - ZLK$ is a lower triangular block matrix with its leading diagonal blocks equal to identity matrices. It follows that $|\, C\,| = 1$ and from (4.108) that $|\,\Omega\,|^{-1} = |\,\Omega^{-1}\,| = |\,C'F^{-1}C\,| = |\,C\,|\cdot|\,F\,|^{-1}\cdot|\,C\,| = |\,F\,|^{-1}$. This result is particularly useful for the evaluation of the log of the density $p(Y_n)$ in (4.103). By applying the Cholesky decomposition to (4.103), we obtain

$$\log p(Y_n) = \text{constant} - \frac{1}{2}\log|\,F\,| - \frac{1}{2}v'F^{-1}v, \tag{4.109}$$

which follows directly from (4.107) and (4.108). The Kalman filter computes v and F in a computationally efficient way and is therefore instrumental for the evaluation of (4.109).

4.13.4 Smoothing in matrix form

Let $\hat{\varepsilon} = (\hat{\varepsilon}_1', \dots, \hat{\varepsilon}_n')'$ where $\hat{\varepsilon}_t = \mathrm{E}(\varepsilon_t|Y_n)$, for $t = 1, \dots, n$, is evaluated as described in Subsection 4.5.3. The smoothed observation disturbance vector $\hat{\varepsilon}$ can be obtained directly via the application of Lemma 1, that is

$$\hat{\varepsilon} = \mathrm{E}(\varepsilon|Y_n) = \mathrm{Cov}(\varepsilon, Y_n)\Omega^{-1}(Y_n - \mu).$$

Since $\mathrm{Cov}(\varepsilon, Y_n) = H$, it follows by the substitution of (4.107) and (4.108) that

$$\begin{aligned} \hat{\varepsilon} &= H\Omega^{-1}(Y_n - \mu) \\ &= Hu, \end{aligned}$$

where

$$\begin{aligned} u &= \Omega^{-1}(Y_n - \mu) \\ &= C'F^{-1}v \\ &= (I - K'L'Z')F^{-1}v \\ &= F^{-1}v - K'r, \end{aligned}$$

with $r = L'Z'F^{-1}v$. It is easily verified that the definitions of u and r are consistent with the definitions of their elements in (4.59) and (4.38), respectively.

Let $\hat{\eta} = (\hat{\eta}_1', \dots, \hat{\eta}_n')'$ where $\hat{\eta}_t = \mathrm{E}(\eta_t|Y_n)$, for $t = 1, \dots, n$, is evaluated as described in Subsection 4.5.3. We obtain the stack of smoothed state disturbance vector $\hat{\eta}$ directly via

$$\begin{aligned} \hat{\eta} &= \mathrm{Cov}(\eta, Y_n)\Omega^{-1}(Y_n - \mu) \\ &= QR'T'Z'u \\ &= QR'r, \end{aligned}$$

where $r = T'Z'u = T'Z'\Omega^{-1}(Y_n - \mu) = L'Z'F^{-1}v$ since $CZT = ZL$. This result is consistent with the definitions of the elements of $\hat{\eta}$, that is $\hat{\eta}_t = Q_t R_t' r_t$ where r_t is evaluated by $r_{t-1} = Z_t' u_t + T_t' r_t$; see Subsection 4.5.3.

Finally, we obtain the smoothed estimator of α via

$$\begin{aligned}\hat{\alpha} &= \mathrm{E}(\alpha) + \mathrm{Cov}(\alpha, Y_n)\Omega^{-1}(Y_n - \mu) \\ &= \mathrm{E}(\alpha) + \mathrm{Cov}(\alpha, Y_n)u \\ &= Ta_1^* + TQ^*T'Z'u \\ &= Ta_1^* + TQ^*r, \end{aligned} \tag{4.110}$$

since $Y_n = Z\alpha + \varepsilon$ and $\mathrm{Cov}(\alpha, Y_n) = \mathrm{Var}(\alpha)Z' = TQ^*T'Z'$. This is consistent with the way $\hat{\alpha}_t$ is evaluated using fast state smoothing as described in Subsection 4.6.2.

4.13.5 Matrix expressions for signal

Given equation (4.102) and the definition of the signal $\theta = Z\alpha$, we further define

$$\mu = \mathrm{E}(\theta) = \mathrm{E}(Z\alpha) = Za^* = ZTa_1^*, \qquad \Psi = \mathrm{Var}(\theta) = ZV^*Z' = ZTQ^*T'Z'.$$

The logdensity of the signal is therefore given by

$$\log p(\theta) = \text{constant} - \frac{1}{2}\log|\Psi| - \frac{1}{2}(\theta - \mu)'\Psi^{-1}(\theta - \mu). \tag{4.111}$$

Also from equation (4.102) we have

$$\mathrm{E}(Y_n) = \mu, \qquad \mathrm{Var}(Y_n) = \Omega = \Psi + H.$$

Since $\mathrm{Cov}(\theta, Y_n) = \mathrm{Var}(\theta) = \Psi$, it follows from Lemma 1 that the conditional (smoothed) mean and variance of the signal is given by

$$\hat{\theta} = \mathrm{E}(\theta|Y_n) = \mu + \Psi\Omega^{-1}(Y_n - \mu), \qquad \mathrm{Var}(\theta|Y_n) = \Psi - \Psi\Omega^{-1}\Psi. \tag{4.112}$$

In the case of the smoothed mean $\hat{\theta}$, after some matrix manupulation, we obtain

$$\hat{\theta} = (\Psi^{-1} + H^{-1})^{-1}(\Psi^{-1}\mu + H^{-1}Y_n).$$

It should be kept in mind that this expression is computed by the Kalman filter and smoothing recursions. In particular, the application of Kalman filter and disturbance smoother is sufficient since $\hat{\theta} = Y_n - \hat{\varepsilon}$. In the linear Gaussian model (4.12) all random variables are Gaussian and all relations between the variables are linear. Therefore, the mean and the mode are equal. It follows that $\hat{\theta}$ is also the mode of the smoothed logdensity $p(\theta|Y_n)$.

The expression for the smoothed signal $\hat{\theta} = \mathrm{E}(\theta|Y_n)$ leads to an interesting result for the computation of u. It follows from (4.112) that

$$u = \Omega^{-1}(Y_n - \mu) = \Psi^{-1}(\hat{\theta} - \mu). \tag{4.113}$$

Since $\Omega = \Psi + H$, the expression (4.113) implies that u can also be computed by applying the Kalman filter and smoother to a linear Gaussian state space model for $\hat{\theta}$ without observation noise such that $H = 0$ and $\Omega = \Psi$, that is

$$\hat{\theta}_t = Z_t \alpha_t, \qquad \alpha_{t+1} = T_t \alpha_t + R_t \eta_t, \tag{4.114}$$

for $t = 1, \ldots, n$. For example, result (4.113) implies that $\hat{\alpha} = \mathrm{E}(\alpha|Y_n)$ can be computed from applying the Kalman filter and smoother to model (4.114) since

$$\hat{\alpha} = Ta_1^* + TQ^*T'Z'u = Ta_1^* + TQ^*T'Z'\Psi^{-1}(\hat{\theta} - \mu).$$

For specific applications these results can be exploited to improve computational efficiency.

4.13.6 Simulation smoothing

Given the results discussed in Section 4.9 together with the matrix expressions and results in this section, we can develop convenient expressions for simulation smoothing for signal and state vectors. In the case of simulation smoothing for the signal, it follows from the discussion in Subsection 4.9.2 that we can express the draw from $p(\theta|Y_n)$ by

$$\tilde{\theta} = \theta^+ - \hat{\theta}^+ + \hat{\theta},$$

where

$$\theta^+ \sim p(\theta), \qquad \hat{\theta}^+ = \mathrm{E}(\theta|y^+), \qquad \hat{\theta} = \mathrm{E}(\theta|Y_n),$$

with

$$y^+ = \theta^+ + \varepsilon^+, \qquad \theta^+ = Z\alpha^+ = ZT(\alpha_1^{*+} + R\eta^+),$$

and

$$\varepsilon^+ \sim \mathrm{N}(0, H), \qquad \alpha_1^+ \sim \mathrm{N}(a, P), \qquad \eta^+ \sim \mathrm{N}(0, Q).$$

We note that $\alpha_1^{*+\prime} = (\alpha_1^{+\prime}, 0, \ldots, 0)'$. Simulation smooting reduces the application of a single Kalman filter and smoother since

$$\tilde{\theta} - \theta^+ = \hat{\theta} - \hat{\theta}^+ = [\mu + \Psi\Omega^{-1}(Y_n - \mu)] - [\mu + \Psi\Omega^{-1}(y^+ - \mu)] = \Psi\Omega^{-1}(Y_n - y^+),$$

using (4.112). It follows that

$$\tilde{\theta} = \theta^+ + \Psi\Omega^{-1}(Y_n - y^+).$$

Once α^+, $\theta^+ = Z\alpha^+$ and $y^+ = \theta^+ + \varepsilon^+$ are computed using the relations in the linear Gaussian state space model (4.12), the sample $\tilde{\theta} \sim p(\theta|Y_n)$ is obtained from the Kalman filter and smoother applied to model (4.12) with $a_1 = 0$ and for the 'observations' $y_t - y_t^+$, for $t = 1, \ldots, n$. Similar arguments apply to the computation of $\tilde{\alpha}$, $\tilde{\varepsilon}$ and $\tilde{\eta}$.

4.14 Exercises

4.14.1

Taking the notation of Section 4.2 and

$$\Sigma_* = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma'_{xy} & \Sigma_{yy} \end{bmatrix},$$

verify that

$$\Sigma_* = \begin{bmatrix} I & \Sigma_{xy}\Sigma_{yy}^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma'_{xy} & 0 \\ 0 & \Sigma_{yy} \end{bmatrix} \begin{bmatrix} I & 0 \\ \Sigma_{yy}^{-1}\Sigma'_{xy} & I \end{bmatrix},$$

and hence that

$$\Sigma_*^{-1} = \begin{bmatrix} I & 0 \\ -\Sigma_{yy}^{-1}\Sigma'_{xy} & I \end{bmatrix} \begin{bmatrix} \left(\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma'_{xy}\right)^{-1} & 0 \\ 0 & \Sigma_{yy}^{-1} \end{bmatrix} \begin{bmatrix} I & -\Sigma_{xy}\Sigma_{yy}^{-1} \\ 0 & I \end{bmatrix}.$$

Taking

$$p(x, y) = \text{constant} \times \exp\left[-\frac{1}{2} \begin{pmatrix} x - \mu_x \\ y - \mu_y \end{pmatrix}' \Sigma^{-1} \begin{pmatrix} x - \mu_x \\ y - \mu_y \end{pmatrix} \right],$$

obtain $p(x|y)$ and hence prove Lemma 1. This exercise contains the essentials of the proofs of our Lemma 1 in Anderson and Moore (1979, Example 3.2) and Harvey (1989, Appendix to Chapter 3).

4.14.2

Under the conditions of Lemma 1 and using the expression for Σ^{-1} in Exercise 4.14.1, show that $\widehat{x} = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y)$ is the maximum likelihood estimator of x for given y with asymptotic variance $\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma'_{xy}$.

4.14.3

Suppose that the fixed vector λ is regarded as an estimate of a random vector x whose mean vector is μ. Show that the minimum mean square error matrix $\mathrm{E}\left[(\lambda - x)(\lambda - x)'\right]$ is obtained when $\lambda = \mu$.

4.14.4

In the joint distribution, not necessarily normal, of random vectors x and y, suppose that $\bar{x} = \beta + \gamma y$ is an estimate of x given y with mean square error matrix $\mathrm{MSE}(\bar{x}) = \mathrm{E}\left[(\bar{x} - x)(\bar{x} - x)'\right]$. Using Exercise 4.14.3 and details of the proof of Lemma 2, show that the minimum mean square error matrix is obtained when $\bar{x} = \widehat{x}$ where $\widehat{x}$ is defined by (4.6) with $\mathrm{MSE}(\widehat{x})$ given by (4.7).

4.14.5

Adopt the notation of Lemma 3 and assume that $p(y|x) = \mathrm{N}(Zx, H)$, where Z and H are constant matrices of appropriate dimensions.

(a) Show that

$$p(x|y) = \exp(-\frac{1}{2}Q)$$

where

$$\begin{aligned} Q &= x'(Z'H^{-1}Z + \Sigma_{yy}^{-1})x - 2(y'H^{-1}Z + \mu_y'\Sigma_{yy}^{-1})x + \text{constant} \\ &= (x-m)'C^{-1}(x-m) + \text{constant}, \end{aligned}$$

with

$$C^{-1} = \Sigma_{yy}^{-1} + Z'H^{-1}Z, \qquad m = C(Z'H^{-1}y + \Sigma_{yy}^{-1}\mu_y).$$

(b) Using the matrix identity

$$(\Sigma_{yy}^{-1} + Z'H^{-1}Z)^{-1} = \Sigma_{yy} - \Sigma_{yy}Z'(Z\Sigma_{yy}Z' + H)^{-1}Z\Sigma_{yy},$$

show that the result in (a) proves Lemma 3 for this form of $p(x|y)$.

This exercise contains the essentials of the proof in West and Harrison (1997, §17.3.3) of our Lemma 1.

4.14.6

Derive the Kalman filter equations of Subsection 4.3.1 in case $\mathrm{E}(\varepsilon_t\eta_t') = R_t^*$ and $\mathrm{E}(\varepsilon_t\eta_s') = 0$ where R_t^* is a fixed and known $p \times r$ matrix for $t, s = 1, \ldots, n$ and $t \neq s$.

4.14.7

Given the state space model (4.12) and the results in Sections 4.4 and 4.5, derive recursive expressions for

$$\mathrm{Cov}(\varepsilon_t, \alpha_t|Y_n), \qquad \mathrm{Cov}(\eta_t, \alpha_t|Y_n),$$

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

4.14.8

How would you modify the state space model to carry out fixed-lag smoothing when you want to rely only on the smoothing recursions (4.44) in Subsection 4.4.4? See also Exercise 4.14.7.