CHAPTER 11

State-Space Models and Kalman Filter

The state-space model provides a flexible approach to time series analysis, especially for simplifying maximum-likelihood estimation and handling missing values. In this chapter, we discuss the relationship between the state-space model and the ARIMA model, the Kalman filter algorithm, various smoothing methods, and some applications. We begin with a simple model that shows the basic ideas of the state-space approach to time series analysis before introducing the general state-space model. For demonstrations, we use the model to analyze realized volatility series of asset returns, the time-varying coefficient market models, and the quarterly earnings per share of a company.

There are many books on statistical analysis using the state-space model. Durbin and Koopman (2001) provide a recent treatment of the approach, Kim and Nelson (1999) focus on economic applications and regime switching, and Anderson and Moore (1979) give a nice summary of theory and applications of the approach for engineering and optimal control. Many time series textbooks include the Kalman filter and state-space model. For example, Chan (2002), Shumway and Stoffer (2000), Hamilton (1994), and Harvey (1993) all have chapters on the topic. West and Harrison (1997) provide a Bayesian treatment with emphasis on forecasting, and Kitagawa and Gersch (1996) use a smoothing prior approach.

The derivation of Kalman filter and smoothing algorithms necessarily involves heavy notation. Therefore, Section 11.4 could be dry for readers who are interested mainly in the concept and applications of state-space models and can be skipped on the first read.

Analysis of Financial Time Series, Third Edition, By Ruey S. Tsay

11.1 LOCAL TREND MODEL

Consider the univariate time series y_t satisfying

$$y_t = \mu_t + e_t, \qquad e_t \sim N(0, \sigma_e^2), \tag{11.1}$$

$$\mu_{t+1} = \mu_t + \eta_t, \qquad \eta_t \sim N(0, \sigma_\eta^2), \tag{11.2}$$

where $\{e_t\}$ and $\{\eta_t\}$ are two independent Gaussian white noise series and $t = 1, \ldots, T$. The initial value μ_1 is either given or follows a known distribution, and it is independent of $\{e_t\}$ and $\{\eta_t\}$ for $t > 0$. Here μ_t is a pure *random walk* of Chapter 2 with initial value μ_1, and y_t is an observed version of μ_t with added noise e_t. In the literature, μ_t is referred to as the *trend* of the series, which is not directly observable, and y_t is the observed data with observational noise e_t. The dynamic dependence of y_t is governed by that of μ_t because $\{e_t\}$ is not serially correlated.

The model in Eqs. (11.1) and (11.2) can readily be used to analyze realized volatility of an asset price; see Example 11.1. Here μ_t represents the underlying log volatility of the asset price and y_t is the logarithm of realized volatility. The true log volatility is not directly observed but evolves over time according to a random-walk model. On the other hand, y_t is constructed from high-frequency transactions data and subjected to the influence of market microstructure noises. The standard deviation of e_t denotes the scale used to measure the impact of market microstructure noises.

The model in Eqs. (11.1) and (11.2) is a special *linear Gaussian state-space model*. The variable μ_t is called the *state* of the system at time t and is not directly observed. Equation (11.1) provides the link between the data y_t and the state μ_t and is called the *observation equation* with *measurement error* e_t. Equation (11.2) governs the time evolution of the state variable and is the *state equation* (or *state transition equation*) with innovation η_t. The model is also called a *local-level model* in Durbin and Koopman (2001, Chapter 2), which is a simple case of the *structural time series model* of Harvey (1993).

Relationship to ARIMA Model

If there is no measurement error in Eq. (11.1), that is, $\sigma_e = 0$, then $y_t = \mu_t$, which is an ARIMA(0,1,0) model. If $\sigma_e > 0$, that is, there exist measurement errors, then y_t is an ARIMA(0,1,1) model satisfying

$$(1 - B)y_t = (1 - \theta B)a_t, \tag{11.3}$$

where $\{a_t\}$ is a Gaussian white noise with mean zero and variance σ_a^2. The values of θ and σ_a are determined by σ_e and σ_η. This result can be derived as follows.

From Eq. (11.2), we have

$$(1 - B)\mu_{t+1} = \eta_t \qquad \text{or} \qquad \mu_{t+1} = \frac{1}{1 - B}\eta_t.$$

Using this result, Eq. (11.1) can be written as

$$y_t = \frac{1}{1-B}\eta_{t-1} + e_t.$$

Multiplying by $(1 - B)$, we have

$$(1 - B)y_t = \eta_{t-1} + e_t - e_{t-1}.$$

Let $(1 - B)y_t = w_t$. We have $w_t = \eta_{t-1} + e_t - e_{t-1}$. Under the model assumptions, it is easy to see that (a) w_t is Gaussian, (b) $\text{Var}(w_t) = 2\sigma_e^2 + \sigma_\eta^2$, (c) $\text{Cov}(w_t, w_{t-1}) = -\sigma_e^2$, and (d) $\text{Cov}(w_t, w_{t-j}) = 0$ for $j > 1$. Consequently, w_t follows an MA(1) model and can be written as $w_t = (1 - \theta B)a_t$. By equating the variance and lag-1 autocovariance of $w_t = (1 - \theta B)a_t = \eta_{t-1} + e_t - e_{t-1}$, we have

$$(1 + \theta^2)\sigma_a^2 = 2\sigma_e^2 + \sigma_\eta^2, \tag{11.4}$$

$$\theta\sigma_a^2 = \sigma_e^2. \tag{11.5}$$

For given σ_e^2 and σ_η^2, one considers the ratio of the prior two equations to form a quadratic function of θ. This quadratic form has two solutions so one should select the one that satisfies $|\theta| < 1$. The value of σ_a^2 can then be easily obtained. Thus, the state-space model in Eqs. (11.1) and (11.2) is also an ARIMA(0,1,1) model, which is the simple exponential smoothing model of Chapter 2.

On the other hand, for an ARIMA(0,1,1) model with positive θ, one can use the prior two identities to solve for σ_e^2 and σ_η^2, and obtain a local trend model. If θ is negative, then the model can still be put in a state-space form without the observational error, that is, $\sigma_e = 0$. In fact, as will be seen later, an ARIMA model can be transformed into state-space models in many ways. Thus, the linear state-space model is closely related to the ARIMA model.

In practice, what one observes is the y_t series. Thus, based on the data alone, the decision of using ARIMA models or linear state-space models is not critical. Both model representations have pros and cons. The objective of data analysis, substantive issues, and experience all play a role in choosing a statistical model.

Example 11.1. To illustrate the ideas of the state-space model and Kalman filter, we consider the intradaily realized volatility of Alcoa stock from January 2, 2003, to May 7, 2004, for 340 observations. The daily realized volatility used is the sum of squares of intraday 10-minute log returns measured in percentage. No overnight returns or the first 10-minute intraday returns are used. See Chapter 3 for more information about realized volatility. The series used in the demonstration is the logarithm of the daily realized volatility.

Figure 11.1 shows the time plot of the logarithms of the realized volatility of Alcoa stock from January 2, 2003, to May 7, 2004. The transactions data are obtained from the TAQ database of the NYSE. If ARIMA models are entertained,

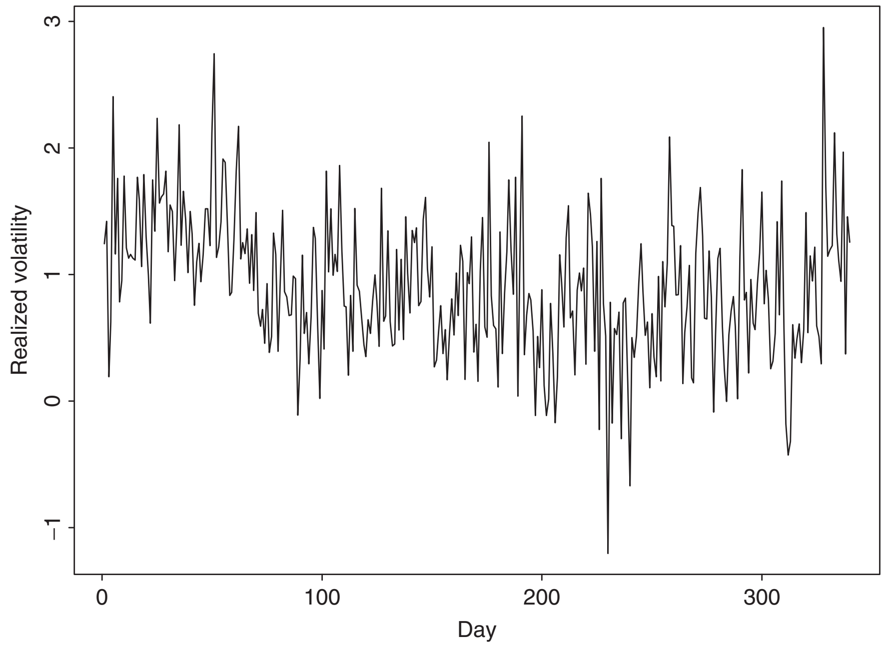


Figure 11.1 Time plot of logarithms of intradaily realized volatility of Alcoa stock from January 2, 2003, to May 7, 2004. Realized volatility is computed from intraday 10-minute log returns measured in percentage.

we obtain an ARIMA(0,1,1) model

$$(1 - B)y_t = (1 - 0.858B)a_t, \qquad \hat{\sigma}_a = 0.5184, \tag{11.6}$$

where y_t is the log realized volatility, and the standard error of $\hat{\theta}$ is 0.028. The residuals show $Q(12) = 12.4$ with a p value of 0.33, indicating that there is no significant serial correlation in the residuals. Similarly, the squared residuals give $Q(12) = 8.2$ with a p value of 0.77, suggesting no ARCH effects in the series.

Since $\hat{\theta}$ is positive, we can transform the ARIMA(0,1,1) model into a local trend model in Eqs. (11.1) and (11.2). The maximum-likelihood estimates (MLE) of the two parameters are $\hat{\sigma}_\eta = 0.0735$ and $\hat{\sigma}_e = 0.4803$. The measurement errors have a larger variance than the state innovations, confirming that intraday high-frequency returns are subject to measurement errors. Details of estimation will be discussed in Section 11.1.7. Here we treat the two estimates as given and use the model to demonstrate application of the Kalman filter. Note that using the model in Eq. (11.6) and the relation in Eqs. (11.4) and (11.5), we obtain $\sigma_e = 0.480$ and $\sigma_\eta = 0.0736$. These values are close to the MLE shown above.

11.1.1 Statistical Inference

Return to the state-space model in Eqs. (11.1) and (11.2). The aim of the analysis is to infer properties of the state μ_t from the data $\{y_t|t = 1, \ldots, T\}$ and the model. Three types of inference are commonly discussed in the literature. They are *filtering*, *prediction*, and *smoothing*. Let $F_t = \{y_1, \ldots, y_t\}$ be the information available at time t (inclusive) and assume that the model is known, including all parameters. The three types of inference can briefly be described as follows:

- *Filtering.* Filtering means to recover the state variable μ_t given F_t, that is, to remove the measurement errors from the data.
- *Prediction.* Prediction means to forecast μ_{t+h} or y_{t+h} for $h > 0$ given F_t, where t is the forecast origin.
- *Smoothing.* Smoothing is to estimate μ_t given F_T, where $T > t$.

A simple analogy of the three types of inference is reading a handwritten note. Filtering is figuring out the word you are reading based on knowledge accumulated from the beginning of the note, predicting is to guess the next word, and smoothing is deciphering a particular word once you have read through the note.

To describe the inference more precisely, we introduce some notation. Let $\mu_{t|j} = E(\mu_t|F_j)$ and $\Sigma_{t|j} = \text{Var}(\mu_t|F_j)$ be, respectively, the conditional mean and variance of μ_t given F_j. Similarly, $y_{t|j}$ denotes the conditional mean of y_t given F_j. Furthermore, let $v_t = y_t - y_{t|t-1}$ and $V_t = \text{Var}(v_t|F_{t-1})$ be the 1-step-ahead forecast error and its variance of y_t given F_{t-1}. Note that the forecast error v_t is independent of F_{t-1} so that the conditional variance is the same as the unconditional variance; that is, $\text{Var}(v_t|F_{t-1}) = \text{Var}(v_t)$. From Eq. (11.1),

$$y_{t|t-1} = E(y_t|F_{t-1}) = E(\mu_t + e_t|F_{t-1}) = E(\mu_t|F_{t-1}) = \mu_{t|t-1}.$$

Consequently,

$$v_t = y_t - y_{t|t-1} = y_t - \mu_{t|t-1} \tag{11.7}$$

and

$$\begin{aligned} V_t &= \text{Var}(y_t - \mu_{t|t-1}|F_{t-1}) = \text{Var}(\mu_t + e_t - \mu_{t|t-1}|F_{t-1}) \\ &= \text{Var}(\mu_t - \mu_{t|t-1}|F_{t-1}) + \text{Var}(e_t|F_{t-1}) = \Sigma_{t|t-1} + \sigma_e^2. \end{aligned} \tag{11.8}$$

It is also easy to see that

$$E(v_t) = E[E(v_t|F_{t-1})] = E[E(y_t - y_{t|t-1}|F_{t-1})] = E[y_{t|t-1} - y_{t|t-1}] = 0,$$

$$\text{Cov}(v_t, y_j) = E(v_t y_j) = E[E(v_t y_j|F_{t-1})] = E[y_j E(v_t|F_{t-1})] = 0, \qquad j < t.$$

Thus, as expected, the 1-step-ahead forecast error is uncorrelated (hence, independent) with y_j for $j < t$. Furthermore, for the linear model in Eqs. (11.1) and (11.2), $\mu_{t|t} = E(\mu_t|F_t) = E(\mu_t|F_{t-1}, v_t)$ and $\Sigma_{t|t} = \mathrm{Var}(\mu_t|F_t) = \mathrm{Var}(\mu_t|F_{t-1}, v_t)$. In other words, the information set F_t can be written as $F_t = \{F_{t-1}, y_t\} = \{F_{t-1}, v_t\}$.

The following properties of multivariate normal distribution are useful in studying the Kalman filter under normality. They can be shown via the multivariate linear regression method or factorization of the joint density. See, also, Appendix B of Chapter 8. For random vectors $\boldsymbol{w}$ and $\boldsymbol{m}$, denote the mean vectors and covariance matrix as $E(\boldsymbol{w}) = \boldsymbol{\mu}_w$, $E(\boldsymbol{m}) = \boldsymbol{\mu}_m$, and $\mathrm{Cov}(\boldsymbol{m}, \boldsymbol{w}) = \boldsymbol{\Sigma}_{mw}$, respectively.

Theorem 11.1. Suppose that $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$ are three random vectors such that their joint distribution is multivariate normal. In addition, assume that the diagonal block covariance matrix $\boldsymbol{\Sigma}_{ww}$ is nonsingular for $w = x, y, z$, and $\boldsymbol{\Sigma}_{yz} = \boldsymbol{0}$. Then,

1. $E(\boldsymbol{x}|\boldsymbol{y}) = \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_y)$.
2. $\mathrm{Var}(\boldsymbol{x}|\boldsymbol{y}) = \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xx}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}$.
3. $E(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{z}) = E(\boldsymbol{x}|\boldsymbol{y}) + \boldsymbol{\Sigma}_{xz}\boldsymbol{\Sigma}_{zz}^{-1}(\boldsymbol{z} - \boldsymbol{\mu}_z)$.
4. $\mathrm{Var}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{z}) = \mathrm{Var}(\boldsymbol{x}|\boldsymbol{y}) - \boldsymbol{\Sigma}_{xz}\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\Sigma}_{zx}$.

11.1.2 Kalman Filter

The goal of the *Kalman filter* is to update knowledge of the state variable recursively when a new data point becomes available. That is, knowing the conditional distribution of μ_t given F_{t-1} and the new data y_t, we would like to obtain the conditional distribution of μ_t given F_t, where, as before, $F_j = \{y_1, \ldots, y_j\}$. Since $F_t = \{F_{t-1}, v_t\}$, giving y_t and F_{t-1} is equivalent to giving v_t and F_{t-1}. Consequently, to derive the Kalman filter, it suffices to consider the joint conditional distribution of $(\mu_t, v_t)'$ given F_{t-1} before applying Theorem 11.1.

The conditional distribution of v_t given F_{t-1} is normal with mean zero and variance given in Eq. (11.8), and that of μ_t given F_{t-1} is also normal with mean $\mu_{t|t-1}$ and variance $\Sigma_{t|t-1}$. Furthermore, the joint distribution of $(\mu_t, v_t)'$ given F_{t-1} is also normal. Thus, what remains to be solved is the conditional covariance between μ_t and v_t given F_{t-1}. From the definition,

$$\begin{aligned}
\mathrm{Cov}(\mu_t, v_t|F_{t-1}) &= E(\mu_t v_t|F_{t-1}) = E[\mu_t(y_t - \mu_{t|t-1})|F_{t-1}] \quad \text{[by Eq. (11.7)]} \\
&= E[\mu_t(\mu_t + e_t - \mu_{t|t-1})|F_{t-1}] \\
&= E[\mu_t(\mu_t - \mu_{t|t-1})|F_{t-1}] + E(\mu_t e_t|F_{t-1}) \\
&= E[(\mu_t - \mu_{t|t-1})^2|F_{t-1}] = \mathrm{Var}(\mu_t|F_{t-1}) = \Sigma_{t|t-1}, \qquad (11.9)
\end{aligned}$$

where we have used the fact that $E[\mu_{t|t-1}(\mu_t - \mu_{t|t-1})|F_{t-1}] = 0$. Putting the results together, we have

$$\begin{bmatrix} \mu_t \\ v_t \end{bmatrix}_{F_{t-1}} \sim N\left(\begin{bmatrix} \mu_{t|t-1} \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{t|t-1} & \Sigma_{t|t-1} \\ \Sigma_{t|t-1} & V_t \end{bmatrix}\right).$$

By Theorem 11.1, the conditional distribution of μ_t given F_t is normal with mean and variance

$$\mu_{t|t} = \mu_{t|t-1} + \frac{\Sigma_{t|t-1} v_t}{V_t} = \mu_{t|t-1} + K_t v_t, \tag{11.10}$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - \frac{\Sigma^2_{t|t-1}}{V_t} = \Sigma_{t|t-1}(1 - K_t), \tag{11.11}$$

where $K_t = \Sigma_{t|t-1}/V_t$ is commonly referred to as the *Kalman gain*, which is the regression coefficient of μ_t on v_t. From Eq. (11.10), Kalman gain is the factor that governs the contribution of the new shock v_t to the state variable μ_t.

Next, one can make use of the knowledge of μ_t given F_t to predict μ_{t+1} via Eq. (11.2). Specifically, we have

$$\mu_{t+1|t} = E(\mu_t + \eta_t | F_t) = E(\mu_t | F_t) = \mu_{t|t}, \tag{11.12}$$

$$\Sigma_{t+1|t} = \mathrm{Var}(\mu_{t+1}|F_t) = \mathrm{Var}(\mu_t|F_t) + \mathrm{Var}(\eta_t) = \Sigma_{t|t} + \sigma^2_\eta. \tag{11.13}$$

Once the new data y_{t+1} is observed, one can repeat the above procedure to update knowledge of μ_{t+1}. This is the famous *Kalman filter* algorithm proposed by Kalman (1960).

In summary, putting Eqs. (11.7) and (11.13) together and conditioning on the initial assumption that μ_1 is distributed as $N(\mu_{1|0}, \Sigma_{1|0})$, the Kalman filter for the local trend model is as follows:

$$\begin{aligned} v_t &= y_t - \mu_{t|t-1}, \\ V_t &= \Sigma_{t|t-1} + \sigma^2_e, \\ K_t &= \Sigma_{t|t-1}/V_t, \\ \mu_{t+1|t} &= \mu_{t|t-1} + K_t v_t, \\ \Sigma_{t+1|t} &= \Sigma_{t|t-1}(1 - K_t) + \sigma^2_\eta, \qquad t = 1, \ldots, T. \end{aligned} \tag{11.14}$$

There are many ways to derive the Kalman filter. We use Theorem 11.1, which describes some properties of multivariate normal distribution, for its simplicity. In practice, the choice of initial values $\Sigma_{1|0}$ and $\mu_{1|0}$ requires some attention and we shall discuss it later in Section 11.1.6. For the local trend model in Eqs. (11.1) and (11.2), the two parameters σ_e and σ_η can be estimated via the maximum-likelihood method. Again, the Kalman filter is useful in evaluating the likelihood function of the data in estimation. We shall discuss estimation in Section 11.1.7.

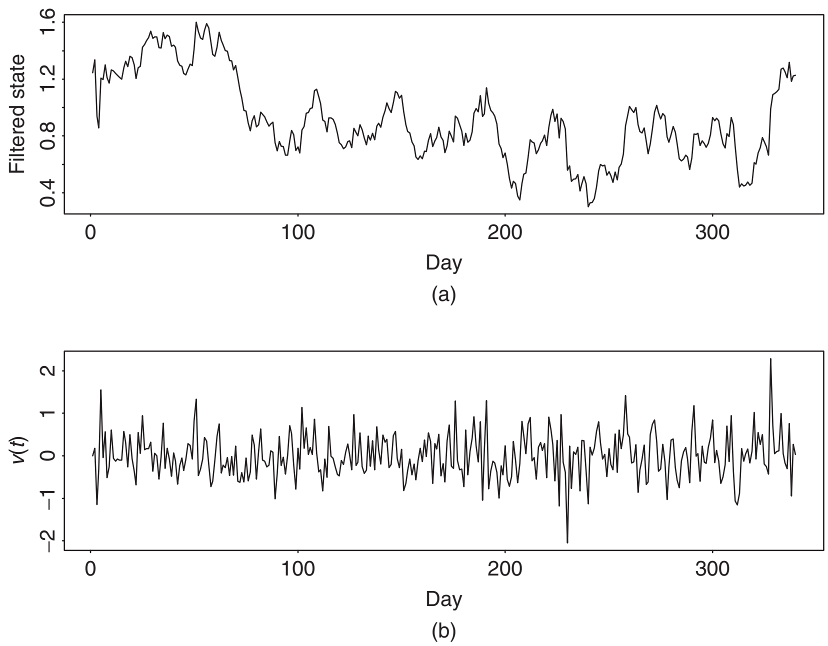


Figure 11.2 Time plots of output of Kalman filter applied to daily realized log volatility of Alcoa stock based on local trend state-space model: (a) filtered state $\mu_{t|t}$ and (b) 1-step-ahead forecast error v_t.

Example 11.1 (Continued). To illustrate application of the Kalman filter, we use the fitted state-space model for daily realized volatility of Alcoa stock returns and apply the Kalman filter algorithm to the data with $\Sigma_{1|0} = \infty$ and $\mu_{1|0} = 0$. The choice of these initial values will be discussed in Section 11.1.6. Figure 11.2(a) shows the time plot of the filtered state variable $\mu_{t|t}$, and Figure 11.2(b) is the time plot of the 1-step-ahead forecast error v_t. Compared with Figure 11.1, the filtered states are smoother. The forecast errors appear to be stable and center around zero. These forecast errors are out-of-sample 1-step-ahead prediction errors.

11.1.3 Properties of Forecast Error

The 1-step-ahead forecast errors $\{v_t\}$ are useful in many applications, hence it pays to study carefully their properties. Given the initial values $\Sigma_{1|0}$ and $\mu_{1|0}$, which are independent of y_t, the Kalman filter enables us to compute v_t recursively as a linear function of $\{y_1, \ldots, y_t\}$. Specifically, by repeated substitutions,

$$v_1 = y_1 - \mu_{1|0},$$

$$v_2 = y_2 - \mu_{2|1} = y_2 - \mu_{1|0} - K_1(y_1 - \mu_{1|0}),$$

$$v_3 = y_3 - \mu_{3|2} = y_3 - \mu_{1|0} - K_2(y_2 - \mu_{1|0}) - K_1(1 - K_2)(y_1 - \mu_{1|0}),$$

and so on. This transformation can be written in matrix form as

$$\boldsymbol{v} = \boldsymbol{K}(\boldsymbol{y} - \mu_{1|0}\mathbf{1}_T), \tag{11.15}$$

where $\boldsymbol{v} = (v_1, \ldots, v_T)'$, $\boldsymbol{y} = (y_1, \ldots, y_T)'$, $\mathbf{1}_T$ is the T-dimensional vector of ones, and $\boldsymbol{K}$ is a lower triangular matrix defined as

$$\boldsymbol{K} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ k_{21} & 1 & 0 & \cdots & 0 \\ k_{31} & k_{32} & 1 & & 0 \\ \vdots & \vdots & & & \vdots \\ k_{T1} & k_{T2} & k_{T3} & \cdots & 1 \end{bmatrix},$$

where $k_{i,i-1} = -K_{i-1}$ and $k_{ij} = -(1-K_{i-1})(1-K_{i-2})\cdots(1-K_{j+1})K_j$ for $i = 2, \ldots, T$ and $j = 1, \ldots, i-2$. It should be noted that, from the definition, the Kalman gain K_t does not depend on $\mu_{1|0}$ or the data $\{y_1, \ldots, y_t\}$; it depends on $\Sigma_{1|0}$ and σ_e^2 and σ_η^2.

The transformation in Eq. (11.5) has several important implications. First, $\{v_t\}$ are mutually independent under the normality assumption. To show this, consider the joint probability density function of the data

$$p(y_1, \ldots, y_T) = p(y_1)\prod_{j=2}^{T} p(y_j|F_{j-1}).$$

Equation (11.15) indicates that the transformation from y_t to v_t has a unit Jacobian so that $p(\boldsymbol{v}) = p(\boldsymbol{y})$. Furthermore, since $\mu_{1|0}$ is given, $p(v_1) = p(y_1)$. Consequently, the joint probability density function of $\boldsymbol{v}$ is

$$p(\boldsymbol{v}) = p(\boldsymbol{y}) = p(y_1)\prod_{j=2}^{T} p(y_j|F_{j-1}) = p(v_1)\prod_{j}^{T} p(v_j) = \prod_{j=1}^{T} p(v_j).$$

This shows that $\{v_t\}$ are mutually independent.

Second, the Kalman filter provides a Cholesky decomposition of the covariance matrix of $\boldsymbol{y}$. To see this, let $\boldsymbol{\Omega} = \mathrm{Cov}(\boldsymbol{y})$. Equation (11.15) shows that $\mathrm{Cov}(\boldsymbol{v}) = \boldsymbol{K}\boldsymbol{\Omega}\boldsymbol{K}'$. On the other hand, $\{v_t\}$ are mutually independent with $\mathrm{Var}(v_t) = V_t$. Therefore, $\boldsymbol{K}\boldsymbol{\Omega}\boldsymbol{K}' = \mathrm{diag}\{V_1, \ldots, V_T\}$, which is precisely a Cholesky decomposition of $\boldsymbol{\Omega}$. The elements k_{ij} of the matrix $\boldsymbol{K}$ thus have some nice interpretations; see Chapter 10.

State Error Recursion

Turn to the estimation error of the state variable μ_t. Define

$$x_t = \mu_t - \mu_{t|t-1}$$

as the forecast error of the state variable μ_t given data F_{t-1}. From Section 11.1.1, $\mathrm{Var}(x_t|F_{t-1}) = \Sigma_{t|t-1}$. From the Kalman filter in Eq. (11.14),

$$v_t = y_t - \mu_{t|t-1} = \mu_t + e_t - \mu_{t|t-1} = x_t + e_t,$$

and

$$\begin{aligned} x_{t+1} &= \mu_{t+1} - \mu_{t+1|t} = \mu_t + \eta_t - (\mu_{t|t-1} + K_t v_t) \\ &= x_t + \eta_t - K_t v_t = x_t + \eta_t - K_t(x_t + e_t) = L_t x_t + \eta_t - K_t e_t, \end{aligned}$$

where $L_t = 1 - K_t = 1 - \Sigma_{t|t-1}/V_t = (V_t - \Sigma_{t|t-1})/V_t = \sigma_e^2/V_t$. Consequently, for the state errors, we have

$$v_t = x_t + e_t, \qquad x_{t+1} = L_t x_t + \eta_t - K_t e_t, \qquad t = 1, \ldots, T, \tag{11.16}$$

where $x_1 = \mu_1 - \mu_{1|0}$. Equation (11.16) is in the form of a time-varying state-space model with x_t being the state variable and v_t the observation.

11.1.4 State Smoothing

Next we consider the estimation of the state variables $\{\mu_1, \ldots, \mu_T\}$ given the data F_T and the model. That is, given the state-space model in Eqs. (11.1) and (11.2), we wish to obtain the conditional distribution $\mu_t|F_T$ for all t. To this end, we first recall some facts available about the model:

- All distributions involved are normal so that we can write the conditional distribution of μ_t given F_T as $N(\mu_{t|T}, \Sigma_{t|T})$, where $t \le T$. We refer to $\mu_{t|T}$ as the *smoothed state* at time t and $\Sigma_{t|T}$ as the *smoothed state variance*.
- Based on the properties of $\{v_t\}$ shown in Section 11.1.3, $\{v_1, \ldots, v_T\}$ are mutually independent and are linear functions of $\{y_1, \ldots, y_T\}$.
- If $y_1, \ldots, y_T$ are fixed, then F_{t-1} and $\{v_t, \ldots, v_T\}$ are fixed, and vice versa.
- $\{v_t, \ldots, v_T\}$ are independent of F_{t-1} with mean zero and variance $\mathrm{Var}(v_j) = V_j$ for $j \ge t$.

Applying Theorem 11.1(3) to the conditional joint distribution of $(\mu_t, v_t, \ldots, v_T)$ given F_{t-1}, we have

$$\begin{aligned} \mu_{t|T} &= E(\mu_t|F_T) = E(\mu_t|F_{t-1}, v_t, \ldots, v_T) \\ &= E(\mu_t|F_{t-1}) + \mathrm{Cov}[\mu_t, (v_t, \ldots, v_T)']\mathrm{Cov}[(v_t, \ldots, v_T)']^{-1}(v_t, \ldots, v_T)' \\ &= \mu_{t|t-1} + \begin{bmatrix} \mathrm{Cov}(\mu_t, v_t) \\ \mathrm{Cov}(\mu_t, v_{t+1}) \\ \vdots \\ \mathrm{Cov}(\mu_t, v_T) \end{bmatrix}' \begin{bmatrix} V_t & 0 & \cdots & 0 \\ 0 & V_{t+1} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & V_T \end{bmatrix}^{-1} \begin{bmatrix} v_t \\ v_{t+1} \\ \vdots \\ v_T \end{bmatrix} \end{aligned}$$

$$= \mu_{t|t-1} + \sum_{j=t}^{T} \text{Cov}(\mu_t, v_j) V_j^{-1} v_j. \tag{11.17}$$

From the definition and independence of $\{v_t\}$, $\text{Cov}(\mu_t, v_j) = \text{Cov}(x_t, v_j)$ for $j = t, \ldots, T$, and

$$\text{Cov}(x_t, v_t) = E[x_t(x_t + e_t)] = \text{Var}(x_t) = \Sigma_{t|t-1},$$

$$\text{Cov}(x_t, v_{t+1}) = E[x_t(x_{t+1} + e_{t+1})] = E[x_t(L_t x_t + \eta_t - K_t e_t)] = \Sigma_{t|t-1} L_t.$$

Similarly, we have

$$\begin{aligned}
\text{Cov}(x_t, v_{t+2}) &= E[x_t(x_{t+2} + e_{t+2})] = \cdots = \Sigma_{t|t-1} L_t L_{t+1}, \\
&\vdots \\
\text{Cov}(x_t, v_T) &= E[x_t(x_T + e_T)] = \cdots = \Sigma_{t|t-1} \prod_{j=t}^{T-1} L_j.
\end{aligned}$$

Consequently, Eq. (11.17) becomes

$$\begin{aligned}
\mu_{t|T} &= \mu_{t|t-1} + \Sigma_{t|t-1} \frac{v_t}{V_t} + \Sigma_{t|t-1} L_t \frac{v_{t+1}}{V_{t+1}} + \Sigma_{t|t-1} L_t L_{t+1} \frac{v_{t+2}}{V_{t+2}} + \cdots \\
&\equiv \mu_{t|t-1} + \Sigma_{t|t-1} q_{t-1},
\end{aligned}$$

where

$$q_{t-1} = \frac{v_t}{V_t} + L_t \frac{v_{t+1}}{V_{t+1}} + L_t L_{t+1} \frac{v_{t+2}}{V_{t+2}} + \cdots + \left(\prod_{j=t}^{T-1} L_j \right) \frac{v_T}{V_T} \tag{11.18}$$

is a weighted linear combination of the innovations $\{v_t, \ldots, v_T\}$. This weighted sum satisfies

$$\begin{aligned}
q_{t-1} &= \frac{v_t}{V_t} + L_t \left[\frac{v_{t+1}}{V_{t+1}} + L_{t+1} \frac{v_{t+2}}{V_{t+2}} + \cdots + \left(\prod_{j=t+1}^{T-1} L_j \right) \frac{v_T}{V_T} \right] \\
&= \frac{v_t}{V_t} + L_t q_t.
\end{aligned}$$

Therefore, using the initial value $q_T = 0$, we have the backward recursion

$$q_{t-1} = \frac{v_t}{V_t} + L_t q_t, \qquad t = T, T-1, \ldots, 1. \tag{11.19}$$

Putting Eqs. (11.17) and (11.19) together, we have a backward recursive algorithm to compute the smoothed state variables:

$$q_{t-1} = V_t^{-1} v_t + L_t q_t, \qquad \mu_{t|T} = \mu_{t|t-1} + \Sigma_{t|t-1} q_{t-1}, \qquad t = T, \ldots, 1, \tag{11.20}$$

where $q_T = 0$, and $\mu_{t|t-1}$, $\Sigma_{t|t-1}$ and L_t are available from the Kalman filter in Eq. (11.14).

Smoothed State Variance

The variance of the smoothed state variable $\mu_{t|T}$ can be derived in a similar manner via Theorem 11.1(4). Specifically, letting $\boldsymbol{v}_t^T = (v_t, \ldots, v_T)'$, we have

$$\begin{aligned}
\Sigma_{t|T} &= \mathrm{Var}(\mu_t|F_T) = \mathrm{Var}(\mu_t|F_{t-1}, v_t, \ldots, v_T) \\
&= \mathrm{Var}(\mu_t|F_{t-1}) - \mathrm{Cov}[\mu_t, (\boldsymbol{v}_t^T)']\mathrm{Cov}[(\boldsymbol{v}_t^T)]^{-1}\mathrm{Cov}[\mu_t, (\boldsymbol{v}_t^T)] \\
&= \Sigma_{t|t-1} - \sum_{j=t}^{T} [\mathrm{Cov}(\mu_t, v_j)]^2 V_j^{-1},
\end{aligned} \tag{11.21}$$

where $\mathrm{Cov}(\mu_t, v_j) = \mathrm{Cov}(x_t, v_j)$ are given earlier after Eq. (11.17). Thus,

$$\begin{aligned}
\Sigma_{t|T} &= \Sigma_{t|t-1} - \Sigma_{t|t-1}^2 \frac{1}{V_t} - \Sigma_{t|t-1}^2 L_t^2 \frac{1}{V_{t+1}} - \cdots - \Sigma_{t|t-1}^2 \left(\prod_{j=t}^{T-1} L_j^2 \right) \frac{1}{V_T} \\
&\equiv \Sigma_{t|t-1} - \Sigma_{t|t-1}^2 M_{t-1},
\end{aligned} \tag{11.22}$$

where

$$M_{t-1} = \frac{1}{V_t} + L_t^2 \frac{1}{V_{t+1}} + L_t^2 L_{t+1}^2 \frac{1}{V_{t+2}} + \cdots + \left(\prod_{j=t}^{T-1} L_j^2 \right) \frac{1}{V_T},$$

is a weighted linear combination of the inverses of variances of the 1-step-ahead forecast errors after time $t-1$. Let $M_T = 0$ because no 1-step-ahead forecast error is available after time index T. The statistic M_{t-1} can be written as

$$\begin{aligned}
M_{t-1} &= \frac{1}{V_t} + L_t^2 \left[\frac{1}{V_{t+1}} + L_{t+1}^2 \frac{1}{V_{t+2}} + \cdots + \left(\prod_{j=t+1}^{T-1} L_j^2 \right) \frac{1}{V_T} \right] \\
&= \frac{1}{V_t} + L_t^2 M_t, \qquad t = T, T-1, \ldots, 1.
\end{aligned}$$

Note that from the independence of $\{v_t\}$ and Eq. (11.18), we have

$$\mathrm{Var}(q_{t-1}) = \frac{1}{V_t} + L_t^2 \frac{1}{V_{t+1}} + \cdots + \left(\prod_{j=t}^{T-1} L_j^2 \right) \frac{1}{V_T} = M_{t-1}.$$

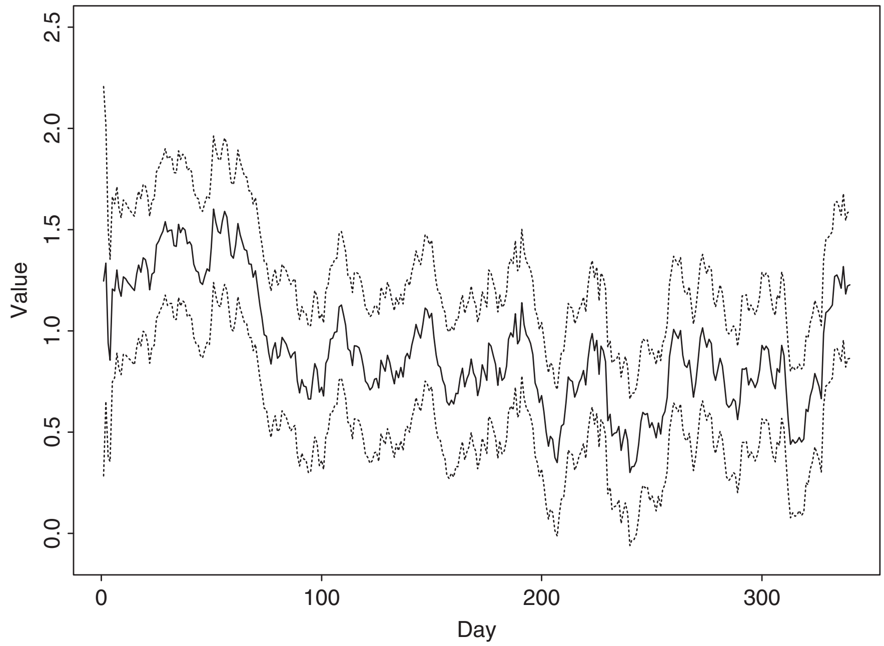


Figure 11.3 Filtered state variable $\mu_{t|t}$ and its 95% pointwise confidence interval for daily log realized volatility of Alcoa stock returns based on fitted local-trend state-space model.

Combining the results, variances of the smoothed state variables can be computed efficiently via the backward recursion

$$M_{t-1} = V_t^{-1} + L_t^2 M_t, \qquad \Sigma_{t|T} = \Sigma_{t|t-1} - \Sigma_{t|t-1}^2 M_{t-1}, \qquad t = T, \ldots, 1, \tag{11.23}$$

where $M_T = 0$.

Example 11.1 (Continued). Applying the Kalman filter and state-smoothing algorithms in Eqs. (11.20) and (11.23) to the daily realized volatility of Alcoa stock using the fitted state-space model, we can easily compute the filtered state $\mu_{t|t}$ and the smoothed state $\mu_{t|T}$ and their variances. Figure 11.3 shows the filtered state variable and its 95% pointwise confidence interval, whereas Figure 11.4 provides the time plot of smoothed state variable and its 95% pointwise confidence interval. As expected, the smoothed state variables are smoother than the filtered state variables. The confidence intervals for the smoothed state variables are also narrower than those of the filtered state variables. Note that the width of the 95% confidence interval of $\mu_{1|1}$ depends on the initial value $\Sigma_{1|0}$.

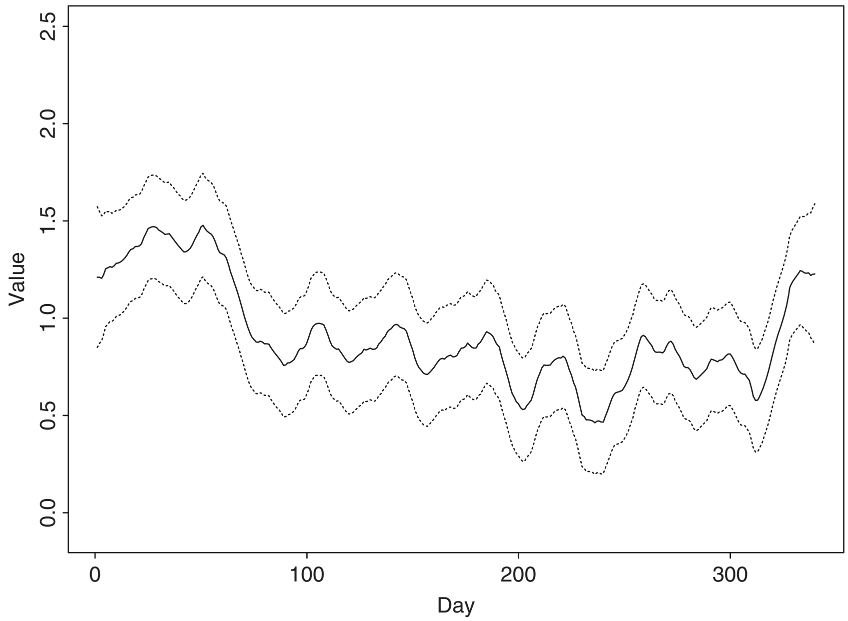


Figure 11.4 Smoothed state variable $\mu_{t|T}$ and its 95% pointwise confidence interval for daily log realized volatility of Alcoa stock returns based on fitted local-trend state-space model.

11.1.5 Missing Values

An advantage of the state-space model is in handling missing values. Suppose that the observations $\{y_t\}_{t=\ell+1}^{\ell+h}$ are missing, where $h \geq 1$ and $1 \leq \ell < T$. There are several ways to handle missing values in state-space formulation. Here we discuss a method that keeps the original time scale and model form. For $t \in \{\ell + 1, \ldots, \ell + h\}$, we can use Eq. (11.2) to express μ_t as a linear combination of $\mu_{\ell+1}$ and $\{\eta_j\}_{j=\ell+1}^{t-1}$. Specifically,

$$\mu_t = \mu_{t-1} + \eta_{t-1} = \cdots = \mu_{\ell+1} + \sum_{j=\ell+1}^{t-1} \eta_j,$$

where it is understood that the summation term is zero if its lower limit is greater than its upper limit. Therefore, for $t \in \{\ell + 1, \ldots, \ell + h\}$,

$$E(\mu_t|F_{t-1}) = E(\mu_t|F_\ell) = \mu_{\ell+1|\ell},$$
$$\text{Var}(\mu_t|F_{t-1}) = \text{Var}(\mu_t|F_\ell) = \Sigma_{\ell+1|\ell} + (t - \ell - 1)\sigma_\eta^2.$$

Consequently, we have

$$\mu_{t|t-1} = \mu_{t-1|t-2}, \qquad \Sigma_{t|t-1} = \Sigma_{t-1|t-2} + \sigma_\eta^2, \tag{11.24}$$

for $t = \ell + 2, \ldots, \ell + h$. These results show that we can continue to apply the Kalman filter algorithm in Eq. (11.14) by taking $v_t = 0$ and $K_t = 0$ for $t = \ell + 1, \ldots, \ell + h$. This is rather natural because when y_t is missing, there is no new innovation or new Kalman gain so that $v_t = 0$ and $K_t = 0$.

11.1.6 Effect of Initialization

In this section, we consider the effects of initial condition $\mu_1 \sim N(\mu_{1|0}, \Sigma_{1|0})$ on the Kalman filter and state smoothing. From the Kalman filter in Eq. (11.14),

$$v_1 = y_1 - \mu_{1|0}, \qquad V_1 = \Sigma_{1|0} + \sigma_e^2,$$

and, by Eqs. (11.10)–(11.13),

$$\mu_{2|1} = \mu_{1|0} + \frac{\Sigma_{1|0}}{V_1} v_1 = \mu_{1|0} + \frac{\Sigma_{1|0}}{\Sigma_{1|0} + \sigma_e^2}(y_1 - \mu_{1|0}),$$

$$\Sigma_{2|1} = \Sigma_{1|0}\left(1 - \frac{\Sigma_{1|0}}{\Sigma_{1|0} + \sigma_e^2}\right) + \sigma_\eta^2 = \frac{\Sigma_{1|0}}{\Sigma_{1|0} + \sigma_e^2}\sigma_e^2 + \sigma_\eta^2.$$

Therefore, letting $\Sigma_{1|0}$ increase to infinity, we have $\mu_{2|1} = y_1$ and $\Sigma_{2|1} = \sigma_e^2 + \sigma_\eta^2$. This is equivalent to treating y_1 as fixed and assuming $\mu_1 \sim N(y_1, \sigma_e^2)$. In the literature, this approach to initializing the Kalman filter is called *diffuse initialization* because a very large $\Sigma_{1|0}$ means one is uncertain about the initial condition.

Next, turn to the effect of diffuse initialization on state smoothing. It is obvious that based on the results of Kalman filtering, state smoothing is not affected by the diffuse initialization for $t = T, \ldots, 2$. Thus, we focus on μ_1 given F_T. From Eq. (11.20) and the definition of $L_1 = 1 - K_1 = V_1^{-1}\sigma_e^2$,

$$\begin{aligned}
\mu_{1|T} &= \mu_{1|0} + \Sigma_{1|0} q_0 \\
&= \mu_{1|0} + \Sigma_{1|0}\left[\frac{1}{\Sigma_{1|0} + \sigma_e^2} v_1 + \left(1 - \frac{\Sigma_{1|0}}{\Sigma_{1|0} + \sigma_e^2}\right) q_1\right] \\
&= \mu_{1|0} + \frac{\Sigma_{1|0}}{\Sigma_{1|0} + \sigma_e^2}(v_1 + \sigma_e^2 q_1).
\end{aligned}$$

Letting $\Sigma_{1|0} \to \infty$, we have $\mu_{1|T} = \mu_{1|0} + v_1 + \sigma_e^2 q_1 = y_1 + \sigma_e^2 q_1$. Furthermore, from Eq. (11.23) and using $V_1 = \Sigma_{1|0} + \sigma_e^2$, we have

$$\begin{aligned}
\Sigma_{1|T} &= \Sigma_{1|0} - \Sigma_{1|0}^2\left[\frac{1}{\Sigma_{1|0} + \sigma_e^2} + \left(1 - \frac{\Sigma_{1|0}}{\Sigma_{1|0} + \sigma_e^2}\right)^2 M_1\right] \\
&= \Sigma_{1|0}\left(1 - \frac{\Sigma_{1|0}}{\Sigma_{1|0} + \sigma_e^2}\right) - \left(1 - \frac{\Sigma_{1|0}}{\Sigma_{1|0} + \sigma_e^2}\right)^2 \Sigma_{1|0}^2 M_1 \\
&= \left(\frac{\Sigma_{1|0}}{\Sigma_{1|0} + \sigma_e^2}\right)\sigma_e^2 - \left(\frac{\Sigma_{1|0}}{\Sigma_{1|0} + \sigma_e^2}\right)^2 \sigma_e^4 M_1.
\end{aligned}$$

Thus, letting $\Sigma_{1|0} \to \infty$, we obtain $\Sigma_{1|T} = \sigma_e^2 - \sigma_e^4 M_1$.

Based on the prior discussion, we suggest using diffuse initialization when little is known about the initial value μ_1. However, it might be hard to justify the use of a random variable with infinite variance in real applications. If necessary, one can treat μ_1 as an additional parameter of the state-space model and estimate it jointly with other parameters. This latter approach is closely related to the exact maximum-likelihood estimation of Chapters 2 and 8.

11.1.7 Estimation

In this section, we consider the estimation of σ_e and σ_η of the local trend model in Eqs. (11.1) and (11.2). Based on properties of forecast errors discussed in Section 11.1.3, the Kalman filter provides an efficient way to evaluate the likelihood function of the data for estimation. Specifically, the likelihood function under normality is

$$\begin{aligned}
p(y_1, \ldots, y_T|\sigma_e, \sigma_\eta) &= p(y_1|\sigma_e, \sigma_\eta) \prod_{t=2}^{T} (y_t|F_{t-1}, \sigma_e, \sigma_\eta) \\
&= p(y_1|\sigma_e, \sigma_\eta) \prod_{t=2}^{T} (v_t|F_{t-1}, \sigma_e, \sigma_\eta),
\end{aligned}$$

where $y_1 \sim N(\mu_{1|0}, V_1)$ and $v_t = (y_t - \mu_{t|t-1}) \sim N(0, V_t)$. Consequently, assuming $\mu_{1|0}$ and $\Sigma_{1|0}$ are known, and taking the logarithms, we have

$$\ln[L(\sigma_e, \sigma_\eta)] = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left[\ln(V_t) + \frac{v_t^2}{V_t}\right], \tag{11.25}$$

which involves v_t and V_t. Therefore, the log-likelihood function, including cases with missing values, can be evaluated recursively via the Kalman filter. Many software packages perform state-space model estimation via a Kalman filter algorithm such as Matlab, RATS, and S-Plus. In this chapter, we use the SsfPack program developed by Koopman, Shephard, and Doornik (1999) and available in S-Plus and OX. Both Ssfpack and OX are free and can be downloaded from their websites.

11.1.8 S-Plus Commands Used

We provide here the SsfPack commands used to perform analysis of the daily realized volatility of Aloca stock returns. Only brief explanations are given. For further details of the commands used, see Durbin and Koopman (2001, Section 6.6). S-Plus uses specific notation to specify a state-space model; see Table 11.1. The notation must be followed closely. In Table 11.2, we give some commands and their functions.

TABLE 11.1 State-Space Form and Notation in S-Plus

State-Space Parameter	S-Plus Name
$\boldsymbol{\delta}$	mDelta
$\boldsymbol{\Phi}$	mPhi
$\boldsymbol{\Omega}$	mOmega
$\boldsymbol{\Sigma}$	mSigma

TABLE 11.2 Some Commands of SsfPack Package

Command	Function
SsfFit	Maximum-likelihood estimation
CheckSsf	Create "Ssf" object in S-Plus
KalmanFil	Perform Kalman filtering
KalmanSmo	Perform state smoothing
SsfMomentEst with task "STFIL"	Compute filtered state and variance
SsfMomentEst with task "STSMO"	Compute smoothed state and variance
SsfCondDens with task "STSMO"	Compute smoothed state without variance

In our analysis, we first perform maximum-likelihood estimation of the state-space model in Eqs. (11.1) and (11.2) to obtain estimates of σ_e and σ_η. The initial values used are $\Sigma_{1|0} = -1$ and $\mu_{1|0} = 0$, where -1 signifies diffuse initialization, that is, $\Sigma_{1|0}$ is very large. We then treat the fitted model as given to perform Kalman filtering and state smoothing.

SsfPack and S-Plus Commands for State-Space Model

```
> da = read.table(file='aa-rv-0304.txt',header=F) % load data
> y = log(da[,1]) % log(RV)
> ltm.start=c(3,1)  % Initial parameter values
> P1 = -1   % Initialization of Kalman filter
> a1 = 0
> ltm.m=function(parm){     % Specify a function for the
+ sigma.eta=parm[1]         % local trend model.
+ sigma.e=parm[2]
+ ssf.m=list(mPhi=as.matrix(c(1,1)),
+ mOmega=diag(c(sigma.eta^2,sigma.e^2)),
+ mSigma=as.matrix(c(P1,a1)))
+ CheckSsf(ssf.m)
+ }
% perform estimation
> ltm.mle=SsfFit(ltm.start,y,"ltm.m",lower=c(0,0),
+ upper=c(100,100))
> ltm.mle$parameters
```

```
[1] 0.07350827 0.48026284
> sigma.eta=ltm.mle$parameters[1]
> sigma.eta
[1] 0.07350827
> sigma.e=ltm.mle$parameters[2]
> sigma.e
[1] 0.4802628
% Specify a state-space model in S-Plus.
> ssf.ltm.list=list(mPhi=as.matrix(c(1,1)),
+ mOmega=diag(c(sigma.eta^2,sigma.e^2)),
+ mSigma=as.matrix(c(P1,a1)))
% check validity of the specified model.
> ssf.ltm=CheckSsf(ssf.ltm.list)
> ssf.ltm
$mPhi:
      [,1]
[1,]     1
[2,]     1
$mOmega:
          [,1]      [,2]
[1,] 0.0054035 0.0000000
[2,] 0.0000000 0.2306524
$mSigma:
     [,1]
[1,]   -1
[2,]    0
$mDelta:
     [,1]
[1,]    0
[2,]    0
$mJPhi:
[1] 0
$mJOmega:
[1] 0
$mJDelta:
[1] 0
$mX:
[1] 0
$cT:
[1] 0
$cX:
[1] 0
$cY:
[1] 1
$cSt:
[1] 1
attr(, "class"):
[1] "ssf"
% Apply Kalman filter
```

```
> KalmanFil.ltm=KalmanFil(y,ssf.ltm,task="STFIL")
> names(KalmanFil.ltm)
 [1] "mOut"         "innov"  "std.innov"  "mGain"    "loglike"
 [6] "loglike.conc" "dVar"   "mEst"       "mOffP"    "task"
[11] "err"          "call"
> par(mfcol=c(2,1))   % Obtain plot
> plot(KalmanFil.ltm$ mEst[,1],xlab='day',
+ ylab='filtered state',type='l')
> title(main='(a) Filtered state variable')
> plot(KalmanFil.ltm$ mOut[,1],xlab='day',
+ ylab='v(t)',type='l')
> title(main='(b) Prediction error')
% Obtain residuals and their variances
> KalmanSmo.ltm=KalmanSmo(KalmanFil.ltm,ssf.ltm)
> names(KalmanSmo.ltm)
[1] "state.residuals"   "response.residuals" "state.variance"
[4] "response.variance" "aux.residuals"      "scores"
[7] "call"
% Filtered states
> FiledEst.ltm=SsfMomentEst(y,ssf.ltm,task="STFIL")
> names(FiledEst.ltm)
[1] "state.moment"    "state.variance"  "response.moment"
[4] "response.variance" "task"
% Smoothed states
> SmoedEst.ltm=SsfMomentEst(y,ssf.ltm,task="STSMO")
> names(SmoedEst.ltm)
[1] "state.moment"    "state.variance"  "response.moment"
[4] "response.variance" "task"
% Obtain plots of filtered and smoothed states with 95% C.I.
> up=FiledEst.ltm$ state.moment+
+ 2*sqrt(FiledEst.ltm$ state.variance)
> lw=FiledEst.ltm$ state.moment-
+ 2*sqrt(FiledEst.ltm$ state.variance)
> par(mfcol=c(1,1))
> plot(FiledEst.ltm$ state.moment,type='l',xlab='day',
+ ylab='value',ylim=c(-0.1,2.5))
> lines(1:340,up,lty=2)
> lines(1:340,lw,lty=2)
> title(main='Filed state variable')
> up=SmoedEst.ltm$ state.moment+
+ 2*sqrt(SmoedEst.ltm$ state.variance)
> lw=SmoedEst.ltm$ state.moment-
+ 2*sqrt(SmoedEst.ltm$ state.variance)
> plot(SmoedEst.ltm$ state.moment,type='l',xlab='day',
+ ylab='value',ylim=c(-0.1,2.5))
> lines(1:340,up,lty=2)
> lines(1:340,lw,lty=2)
> title(main='Smoothed state variable')
% Model checking via standardized residuals
```

```
> resi=KalmanFil.ltm$ mOut[,1]*sqrt(KalmanFil.ltm$ mOut[,3])
> archTest(resi)
> autocorTest(resi)
```

For the daily realized volatility of Alcoa stock returns, the fitted local trend model is adequate based on residual analysis. Specifically, given the parameter estimates, we use the Kalman filter to obtain the 1-step-ahead forecast error v_t and its variance V_t. We then compute the standardized forecast error $\tilde{v}_t = v_t/\sqrt{V_t}$ and check the serial correlations and ARCH effects of $\{\tilde{v}_t\}$. We found that $Q(25) = 23.37(0.56)$ for the standardized forecast errors, and the LM test statistic for ARCH effect is 18.48(0.82) for 25 lags, where the number in parentheses denotes p value.

11.2 LINEAR STATE-SPACE MODELS

We now consider the general state-space model. Many dynamic time series models in economics and finance can be represented in state-space form. Examples include the ARIMA models, dynamic linear models with unobserved components, time-varying regression models, and stochastic volatility models. A general Gaussian linear state-space model assumes the form

$$\boldsymbol{s}_{t+1} = \boldsymbol{d}_t + \boldsymbol{T}_t \boldsymbol{s}_t + \boldsymbol{R}_t \boldsymbol{\eta}_t, \tag{11.26}$$

$$\boldsymbol{y}_t = \boldsymbol{c}_t + \boldsymbol{Z}_t \boldsymbol{s}_t + \boldsymbol{e}_t, \tag{11.27}$$

where $\boldsymbol{s}_t = (s_{1t}, \ldots, s_{mt})'$ is an m-dimensional state vector, $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{kt})'$ is a k-dimensional observation vector, $\boldsymbol{d}_t$ and $\boldsymbol{c}_t$ are m- and k-dimensional deterministic vectors, $\boldsymbol{T}_t$ and $\boldsymbol{Z}_t$ are $m \times m$ and $k \times m$ coefficient matrices, $\boldsymbol{R}_t$ is an $m \times n$ matrix often consisting of a subset of columns of the $m \times m$ identity matrix, and $\{\boldsymbol{\eta}_t\}$ and $\{\boldsymbol{e}_t\}$ are n- and k-dimensional Gaussian white noise series such that

$$\boldsymbol{\eta}_t \sim N(\boldsymbol{0}, \boldsymbol{Q}_t), \qquad \boldsymbol{e}_t \sim N(\boldsymbol{0}, \boldsymbol{H}_t),$$

where $\boldsymbol{Q}_t$ and $\boldsymbol{H}_t$ are positive-definite matrices. We assume that $\{\boldsymbol{e}_t\}$ and $\{\boldsymbol{\eta}_t\}$ are independent, but this condition can be relaxed if necessary. The initial state $\boldsymbol{s}_1$ is $N(\boldsymbol{\mu}_{1|0}, \boldsymbol{\Sigma}_{1|0})$, where $\boldsymbol{\mu}_{1|0}$ and $\boldsymbol{\Sigma}_{1|0}$ are given, and is independent of $\boldsymbol{e}_t$ and $\boldsymbol{\eta}_t$ for $t > 0$.

Equation (11.27) is the *measurement* or *observation* equation that relates the vector of observations $\boldsymbol{y}_t$ to the state vector $\boldsymbol{s}_t$, the explanatory variable $\boldsymbol{c}_t$, and the measurement error $\boldsymbol{e}_t$. Equation (11.26) is the *state* or *transition* equation that describes a first-order Markov Chain to govern the state transition with innovation $\boldsymbol{\eta}_t$. The matrices $\boldsymbol{T}_t$, $\boldsymbol{R}_t$, $\boldsymbol{Q}_t$, $\boldsymbol{Z}_t$, and $\boldsymbol{H}_t$ are known and referred to as *system matrices*. These matrices are often sparse, and they can be functions of some parameters $\boldsymbol{\theta}$, which can be estimated by the maximum-likelihood method.

The state-space model in Eqs. (11.26) and (11.27) can be rewritten in a compact form as

$$\begin{bmatrix} \boldsymbol{s}_{t+1} \\ \boldsymbol{y}_t \end{bmatrix} = \boldsymbol{\delta}_t + \boldsymbol{\Phi}_t \boldsymbol{s}_t + \boldsymbol{u}_t, \tag{11.28}$$

where

$$\boldsymbol{\delta}_t = \begin{bmatrix} \boldsymbol{d}_t \\ \boldsymbol{c}_t \end{bmatrix}, \qquad \boldsymbol{\Phi}_t = \begin{bmatrix} \boldsymbol{T}_t \\ \boldsymbol{Z}_t \end{bmatrix}, \qquad \boldsymbol{u}_t = \begin{bmatrix} \boldsymbol{R}_t \boldsymbol{\eta}_t \\ \boldsymbol{e}_t \end{bmatrix},$$

and $\{\boldsymbol{u}_t\}$ is a sequence of Gaussian white nosies with mean zero and covariance matrix

$$\boldsymbol{\Omega}_t = \mathrm{Cov}(\boldsymbol{u}_t) = \begin{bmatrix} \boldsymbol{R}_t \boldsymbol{Q}_t \boldsymbol{R}_t' & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{H}_t \end{bmatrix}$$

The case of diffuse initialization is achieved by using

$$\boldsymbol{\Sigma}_{1|0} = \boldsymbol{\Sigma}_* + \lambda \boldsymbol{\Sigma}_\infty,$$

where $\boldsymbol{\Sigma}_*$ and $\boldsymbol{\Sigma}_\infty$ are $m \times m$ symmetric positive-definite matrices and λ is a large real number, which can approach infinity. In S-Plus and SsfPack, the notation

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{1|0} \\ \boldsymbol{\mu}_{1|0}' \end{bmatrix}_{(m+1)\times m}$$

is used; see the notation in Table 11.1.

In many applications, the system matrices are time invariant. However, these matrices can be time varying, making the state-space model flexible.

11.3 MODEL TRANSFORMATION

To appreciate the flexibility of the state-space model, we rewrite some well-known econometric and financial models in state-space form.

11.3.1 CAPM with Time-Varying Coefficients

First, consider the capital asset pricing model (CAPM) with time-varying intercept and slope. The model is

$$\begin{aligned} r_t &= \alpha_t + \beta_t r_{M,t} + e_t, \qquad e_t \sim N(0, \sigma_e^2), \qquad (11.29) \\ \alpha_{t+1} &= \alpha_t + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2), \\ \beta_{t+1} &= \beta_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2), \end{aligned}$$

where r_t is the excess return of an asset, $r_{M,t}$ is the excess return of the market, and the innovations $\{e_t, \eta_t, \epsilon_t\}$ are mutually independent. This CAPM allows for time-varying α and β that evolve as a random walk over time. We can easily rewrite the model as

$$\begin{bmatrix} \alpha_{t+1} \\ \beta_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix} + \begin{bmatrix} \eta_t \\ \epsilon_t \end{bmatrix},$$

$$r_t = [1, r_{M,t}] \begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix} + e_t.$$

Thus, the time-varying CAPM is a special case of the state-space model with $\boldsymbol{s}_t = (\alpha_t, \beta_t)'$, $\boldsymbol{T}_t = \boldsymbol{R}_t = \boldsymbol{I}_2$, the 2×2 identity matrix, $\boldsymbol{d}_t = \boldsymbol{0}$, $c_t = 0$, $\boldsymbol{Z}_t = (1, r_{M,t})$, $\boldsymbol{H}_t = \sigma_e^2$, and $\boldsymbol{Q}_t = \text{diag}\{\sigma_\eta^2, \sigma_\epsilon^2\}$. Furthermore, in the form of Eq. (11.28), we have $\boldsymbol{\delta}_t = \boldsymbol{0}$, $\boldsymbol{u}_t = (\eta_t, \epsilon_t, e_t)'$,

$$\boldsymbol{\Phi}_t = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & r_{M,t} \end{bmatrix}, \qquad \boldsymbol{\Omega}_t = \begin{bmatrix} \sigma_\eta^2 & 0 & 0 \\ 0 & \sigma_\epsilon^2 & 0 \\ 0 & 0 & \sigma_e^2 \end{bmatrix}.$$

If diffuse initialization is used, then

$$\boldsymbol{\Sigma} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 0 & 0 \end{bmatrix}.$$

SsfPack/S-Plus Specification of Time-Varying Models

For the CAPM in Eq. (11.29), $\boldsymbol{\Phi}_t$ contains $r_{M,t}$, which is time varying. Some special input is required to specify such a model in SsfPack. Basically, it requires two additional variables: (a) a data matrix $\boldsymbol{X}$ that stores $\boldsymbol{Z}_t$ and (b) an index matrix for $\boldsymbol{\Phi}_t$ that identifies $\boldsymbol{Z}_t$ from the data matrix. The notation for index matrices of the state-space model in Eq. (11.28) is given in Table 11.3. Note that the matrix $\boldsymbol{J}_\Phi$ must have the same dimension as $\boldsymbol{\Phi}_t$. The elements of $\boldsymbol{J}_\Phi$ are all set to -1 except the elements for which the corresponding elements of $\boldsymbol{\Phi}_t$ are time varying. The nonnegative index value of $\boldsymbol{J}_\Phi$ indicates the column of the data matrix $\boldsymbol{X}$, which contains the time-varying values.

To illustrate, consider the monthly simple excess returns of General Motors stock from January 1990 to December 2003 used in Chapter 9. The monthly simple

TABLE 11.3 Notation and Name Used in SsfPack/S-Plus for Time-Varying State-Space Model

Index Matrix	Name Used in SsfPack/S-Plus
$\boldsymbol{J}_\delta$	mJDelta
$\boldsymbol{J}_\Phi$	mJPhi
$\boldsymbol{J}_\Omega$	mJOmega
Time-Varying Data Matrix	Name Used in SsfPack/S-Plus
$\boldsymbol{X}$	mX

excess return of the S&P 500 composite index is used as the market return. The specification of a time-varying CAPM requires values of the variances σ_η^2, σ_ϵ^2, and σ_e^2. Suppose that $(\sigma_\eta, \sigma_\epsilon, \sigma_e) = (0.02, 0.04, 0.1)$. The state-space specification for the CAPM under SsfPack/S-Plus is given below:

```
> X.mtx=cbind(1,sp) % Here ''sp'' is market excess returns.
> Phi.t = rbind(diag(2),rep(0,2))
> Sigma=-Phi.t
> sigma.eta=.02
> sigma.ep=.04
> sigma.e=.1
> Omega=diag(c(sigma.eta^2,sigma.ep^2,sigma.e^2))
> JPhi = matrix(-1,3,2)  % Create a 3-by-2 matrix of -1.
> JPhi[3,1]=1
> JPhi[3,2]=2
> ssf.tv.capm=list(mPhi=Phi.t,
+ mOmega=Omega,
+ mJPhi=JPhi,
+ mSigma=Sigma,
+ mX=X.mtx)
> ssf.tv.capm
$mPhi:
     [,1] [,2]
[1,]    1    0
[2,]    0    1
[3,]    0    0
 $mOmega:
      [,1]   [,2] [,3]
[1,] 4e-04 0.0000 0.00
[2,] 0e+00 0.0016 0.00
[3,] 0e+00 0.0000 0.01
 $mJPhi:
     [,1] [,2]
[1,]   -1   -1
[2,]   -1   -1
[3,]    1    2
 $mSigma:
     [,1] [,2]
[1,]   -1    0
[2,]    0   -1
[3,]    0    0
 $mX:
numeric matrix: 168 rows, 2 columns.
               sp
 [1,] 1 -0.075187
  ...
[168,] 1  0.05002
```

11.3.2 ARMA Models

Consider a zero-mean ARMA(p, q) process y_t of Chapter 2:

$$\phi(B)y_t = \theta(B)a_t, \qquad a_t \sim N(0, \sigma_a^2), \tag{11.30}$$

where $\phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$ and $\theta(B) = 1 - \sum_{j=1}^{q} \theta_j B^j$, and p and q are non-negative integers. There are many ways to transform such an ARMA model into a state-space form. We discuss three methods available in the literature. Let $m = \max(p, q + 1)$ and rewrite the ARMA model in Eq. (11.30) as

$$y_t = \sum_{i=1}^{m} \phi_i y_{t-i} + a_t - \sum_{j=1}^{m-1} \theta_j a_{t-j}, \tag{11.31}$$

where $\phi_i = 0$ for $i > p$ and $\theta_j = 0$ for $j > q$. In particular, $\theta_m = 0$ because $m > q$.

Akaike's Approach

Akaike (1975) defines the state vector $\boldsymbol{s}_t$ as the minimum collection of variables that contains all the information needed to produce forecasts at the forecast origin t. It turns out that, for the ARMA process in Eq. (11.30) with $m = \max(p, q + 1)$, $\boldsymbol{s}_t = (y_{t|t}, y_{t+1|t}, \ldots, y_{t+m-1|t})'$, where $y_{t+j|t} = E(y_{t+j}|F_t)$ is the conditional expectation of y_{t+j} given $F_t = \{y_1, \ldots, y_t\}$. Since $y_{t|t} = y_t$, the first element of $\boldsymbol{s}_t$ is y_t. Thus, the observation equation is

$$y_t = \boldsymbol{Z}\boldsymbol{s}_t, \tag{11.32}$$

where $\boldsymbol{Z} = (1, 0, \ldots, 0)_{1\times m}$. We derive the transition equation in several steps. First, from the definition,

$$s_{1,t+1} = y_{t+1} = y_{t+1|t} + (y_{t+1} - y_{t+1|t}) = s_{2t} + a_{t+1}, \tag{11.33}$$

where s_{it} is the ith element of $\boldsymbol{s}_t$. Next, consider the MA representation of ARMA models given in Chapter 2. That is,

$$y_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots = \sum_{i=0}^{\infty} \psi_i a_{t-i},$$

where $\psi_0 = 1$ and other ψ weights can be obtained by equating coefficients of B^i in $1 + \sum_{i=1}^{\infty} \psi_i B^i = \theta(B)/\phi(B)$. In particular, we have

$$\begin{aligned}
\psi_1 &= \phi_1 - \theta_1, \\
\psi_2 &= \phi_1\psi_1 + \phi_2 - \theta_2, \\
&\;\vdots
\end{aligned}$$

$$\begin{aligned}\psi_{m-1} &= \phi_1\psi_{m-2} + \phi_2\psi_{m-3} + \cdots + \phi_{m-2}\psi_1 + \phi_{m-1} - \theta_{m-1} \\ &= \sum_{i=1}^{m-1} \phi_i \psi_{m-1-i} - \theta_{m-1}.\end{aligned} \tag{11.34}$$

Using the MA representation, we have, for $j > 0$,

$$\begin{aligned}y_{t+j|t} &= E(y_{t+j}|F_t) = E\left(\sum_{i=0}^{\infty} \psi_i a_{t+j-i}|F_t\right) \\ &= \psi_j a_t + \psi_{j+1} a_{t-1} + \psi_{j+2} a_{t-2} + \cdots\end{aligned}$$

and

$$\begin{aligned}y_{t+j|t+1} &= E(y_{t+j}|F_{t+1}) = \psi_{j-1}a_{t+1} + \psi_j a_t + \psi_{j+1}a_{t-1} + \cdots \\ &= \psi_{j-1}a_{t+1} + y_{t+j|t}.\end{aligned}$$

Thus, for $j > 0$, we have

$$y_{t+j|t+1} = y_{t+j|t} + \psi_{j-1}a_{t+1}. \tag{11.35}$$

This result is referred to as the forecast updating formula of ARMA models. It provides a simple way to update the forecast from origin t to origin $t+1$ when y_{t+1} becomes available. The new information of y_{t+1} is contained in the innovation a_{t+1}, and the time-t forecast is revised based on this new information with weight ψ_{j-1} to compute the time-$(t+1)$ forecast.

Finally, from Eq. (11.31) and using $E(a_{t+j}|F_{t+1}) = 0$ for $j > 1$, we have

$$y_{t+m|t+1} = \sum_{i=1}^{m} \phi_i y_{t+m-i|t+1} - \theta_{m-1}a_{t+1}.$$

Taking Eq. (11.35), the prior equation becomes

$$\begin{aligned}y_{t+m|t+1} &= \sum_{i=1}^{m-1} \phi_i (y_{t+m-i|t} + \psi_{m-i-1}a_{t+1}) + \phi_m y_{t|t} - \theta_{m-1}a_{t+1} \\ &= \sum_{i=1}^{m} \phi_i y_{t+m-i|t} + \left(\sum_{i=1}^{m-1} \phi_i \psi_{m-1-i} - \theta_{m-1}\right) a_{t+1} \\ &= \sum_{i=1}^{m} \phi_i y_{t+m-i|t} + \psi_{m-1}a_{t+1},\end{aligned} \tag{11.36}$$

where the last equality uses Eq. (11.34). Combining Eqs. (11.33) and (11.35) for $j = 2, \ldots, m-1$, and (11.36) together, we have

$$\begin{bmatrix} y_{t+1} \\ y_{t+2|t+1} \\ \vdots \\ y_{t+m-1|t+1} \\ y_{t+m|t+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \phi_m & \phi_{m-1} & \phi_{m-2} & \cdots & \phi_1 \end{bmatrix} \begin{bmatrix} y_t \\ y_{t+1|t} \\ \vdots \\ y_{t+m-2|t} \\ y_{t+m-1|t} \end{bmatrix} + \begin{bmatrix} 1 \\ \psi_1 \\ \vdots \\ \psi_{m-2} \\ \psi_{m-1} \end{bmatrix} a_{t+1}. \tag{11.37}$$

Thus, the transition equation of Akaike's approach is

$$\boldsymbol{s}_{t+1} = \boldsymbol{T}\boldsymbol{s}_t + \boldsymbol{R}\eta_t, \qquad \eta_t \sim N(0, \sigma_a^2), \tag{11.38}$$

where $\eta_t = a_{t+1}$, and $\boldsymbol{T}$ and $\boldsymbol{R}$ are the coefficient matrices in Eq. (11.37).

Harvey's Approach

Harvey (1993, Section 4.4) provides a state-space form with an m-dimensional state vector $\boldsymbol{s}_t$, the first element of which is y_t, that is, $s_{1t} = y_t$. The other elements of $\boldsymbol{s}_t$ are obtained recursively. From the ARMA$(m, m-1)$ model, we have

$$\begin{aligned} y_{t+1} &= \phi_1 y_t + \sum_{i=2}^{m} \phi_i y_{t+1-i} - \sum_{j=1}^{m-1} \theta_j a_{t+1-j} + a_{t+1} \\ &\equiv \phi_1 s_{1t} + s_{2t} + \eta_t, \end{aligned}$$

where $s_{2t} = \sum_{i=2}^{m} \phi_i y_{t+1-i} - \sum_{j=1}^{m-1} \theta_j a_{t+1-j}$, $\eta_t = a_{t+1}$, and as defined earlier $s_{1t} = y_t$. Focusing on $s_{2,t+1}$, we have

$$\begin{aligned} s_{2,t+1} &= \sum_{i=2}^{m} \phi_i y_{t+2-i} - \sum_{j=1}^{m-1} \theta_j a_{t+2-j} \\ &= \phi_2 y_t + \sum_{i=3}^{m} \phi_i y_{t+2-i} - \sum_{j=2}^{m-1} \theta_j a_{t+2-j} - \theta_1 a_{t+1} \\ &\equiv \phi_2 s_{1t} + s_{3t} + (-\theta_1)\eta_t, \end{aligned}$$

where $s_{3t} = \sum_{i=3}^{m} \phi_i y_{t+2-i} - \sum_{j=2}^{m-1} \theta_j a_{t+2-j}$. Next, considering $s_{3,t+1}$, we have

$$\begin{aligned} s_{3,t+1} &= \sum_{i=3}^{m} \phi_i y_{t+3-i} - \sum_{j=2}^{m-1} \theta_j a_{t+3-j} \\ &= \phi_3 y_t + \sum_{i=4}^{m} \phi_i y_{t+3-i} - \sum_{j=3}^{m-1} \theta_j a_{t+3-j} + (-\theta_2) a_{t+1} \\ &\equiv \phi_3 s_{1t} + s_{4t} + (-\theta_2)\eta_t, \end{aligned}$$

where $s_{4t} = \sum_{i=4}^{m} \phi_i y_{t+3-i} - \sum_{j=3}^{m-1} \theta_j a_{t+3-j}$. Repeating the procedure, we have $s_{mt} = \sum_{i=m}^{m} \phi_i y_{t+m-1-i} - \sum_{j=m-1}^{m-1} \theta_j a_{t+m-1-j} = \phi_m y_{t-1} - \theta_{m-1} a_t$. Finally,

$$\begin{aligned} s_{m,t+1} &= \phi_m y_t - \theta_{m-1} a_{t+1} \\ &= \phi_m s_{1t} + (-\theta_{m-1})\eta_t. \end{aligned}$$

Putting the prior equations together, we have a state-space form

$$\boldsymbol{s}_{t+1} = \boldsymbol{T}\boldsymbol{s}_t + \boldsymbol{R}\eta_t, \qquad \eta_t \sim N(0, \sigma_a^2), \tag{11.39}$$

$$y_t = \boldsymbol{Z}\boldsymbol{s}_t, \tag{11.40}$$

where the system matrices are time invariant defined as $\boldsymbol{Z} = (1, 0, \ldots, 0)_{1\times m}$,

$$\boldsymbol{T} = \begin{bmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & & 0 \\ \vdots & & & & \vdots \\ \phi_{m-1} & 0 & 0 & & 1 \\ \phi_m & 0 & 0 & \cdots & 0 \end{bmatrix}, \qquad \boldsymbol{R} = \begin{bmatrix} 1 \\ -\theta_1 \\ \vdots \\ -\theta_{m-1} \end{bmatrix},$$

and $\boldsymbol{d}_t$, $\boldsymbol{c}_t$, and $\boldsymbol{H}_t$ are all zero. The model in Eqs. (11.39) and (11.40) has no measurement errors. It has an advantage that the AR and MA coefficients are directly used in the system matrices.

Aoki's Approach

Aoki (1987, Chapter 4) discusses several ways to convert an ARMA model into a state-space form. First, consider the MA model, that is, $y_t = \theta(B)a_t$. In this case, we can simply define $\boldsymbol{s}_t = (a_{t-q}, a_{t-q+2}, \ldots, a_{t-1})'$ and obtain the state-space form

$$\begin{bmatrix} a_{t-q+1} \\ a_{t-q+2} \\ \vdots \\ a_{t-1} \\ a_t \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} a_{t-q} \\ a_{t-q+1} \\ \vdots \\ a_{t-2} \\ a_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} a_t, \tag{11.41}$$

$$y_t = (-\theta_q, -\theta_{q-1}, \ldots, -\theta_1)\boldsymbol{s}_t + a_t.$$

Note that, in this particular case, a_t appears in both state and measurement equations.

Next, consider the AR model, that is, $\phi(B)z_t = a_t$. Aoki (1987) introduces two methods. The first method is a straightfoward one by defining $\boldsymbol{s}_t = (z_{t-p+1}, \ldots, z_t)'$ to obtain

$$\begin{bmatrix} z_{t-p+2} \\ z_{t-p+3} \\ \vdots \\ z_t \\ z_{t+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & & 1 \\ \phi_p & \phi_{p-1} & \phi_{p-2} & \cdots & \phi_1 \end{bmatrix} \begin{bmatrix} z_{t-p+1} \\ z_{t-p+2} \\ \vdots \\ z_{t+1} \\ z_t \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} a_{t+1}, \tag{11.42}$$

$$z_t = (0, 0, \cdots, 0, 1)\boldsymbol{s}_t.$$

The second method defines the state vector in the same way as the first method except that a_t is removed from the last element; that is, $\boldsymbol{s}_t = z_t - a_t$ if $p = 1$ and $\boldsymbol{s}_t = (z_{t-p+1}, \ldots, z_{t-1}, z_t - a_t)'$ if $p > 1$. Simple algebra shows that

$$\begin{bmatrix} z_{t-p+2} \\ z_{t-p+3} \\ \vdots \\ z_t \\ z_{t+1} - a_{t+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & & 1 \\ \phi_p & \phi_{p-1} & \phi_{p-2} & \cdots & \phi_1 \end{bmatrix} \begin{bmatrix} z_{t-p+1} \\ z_{t-p+2} \\ \vdots \\ z_{t-1} \\ z_t - a_t \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \phi_1 \end{bmatrix} a_t, \tag{11.43}$$

$$z_t = (0, 0, \ldots, 0, 1)\boldsymbol{s}_t + a_t.$$

Again, a_t appears in both transition and measurement equations.

Turn to the ARMA(p, q) model $\phi(B)y_t = \theta(B)a_t$. For simplicity, we assume $q < p$ and introduce an auxiliary variable $z_t = [1/\phi(B)]a_t$. Then, we have

$$\phi(B)z_t = a_t, \qquad y_t = \theta(B)z_t.$$

Since z_t is an AR(p) model, we can use the transition equation in Eq. (11.42) or (11.43). If Eq. (11.42) is used, we can use $y_t = \theta(B)z_t$ to construct the measurement equation as

$$y_t = (-\theta_{p-1}, -\theta_{p-2}, \ldots, -\theta_1, 1)\boldsymbol{s}_t, \tag{11.44}$$

where it is understood that $p > q$ and $\theta_j = 0$ for $j > q$. On the other hand, if Eq. (11.43) is used as the transition equation, we construct the measurement equation as

$$y_t = (-\theta_{p-1}, -\theta_{p-2}, \ldots, -\theta_1, 1)\boldsymbol{s}_t + a_t. \tag{11.45}$$

In summary, there are many state-space representations for an ARMA model. Each representation has its pros and cons. For estimation and forecasting purposes, one can choose any one of those representations. On the other hand, for a time-invariant coefficient state-space model in Eqs. (11.26) and (11.27), one can use the Cayley–Hamilton theorem to show that the observation y_t follows an ARMA(m, m) model, where m is the dimension of the state vector.

SsfPack Command

In SsfPack/S-Plus, a command `GetSsfArma` can be used to transform an ARMA model into a state-space form. Harvey's approach is used. To illustrate, consider the AR(1) model

$$y_t = 0.6y_{t-1} + a_t, \qquad a_t \sim N(0, 0.4^2).$$

The state-space form of the model is

```
> ssf.ar1 = GetSsfArma(ar=0.6,sigma=0.4)
> ssf.ar1
$mPhi:
     [,1]
[1,]  0.6
[2,]  1.0
$mOmega:
     [,1] [,2]
[1,] 0.16    0
[2,] 0.00    0
$mSigma:
     [,1]
[1,] 0.25
[2,] 0.00
```

Since the AR(1) model is stationary, the program uses $\boldsymbol{\Sigma}_{1|0} = \text{Var}(y_t) = (0.4)^2/(1 - 0.6^2) = 0.25$ and $\boldsymbol{\mu}_{1|0} = 0$. These values appear in the matrix `mSigma`.

As a second example, consider the ARMA(2,1) model

$$y_t = 1.2y_{t-1} - 0.35y_{t-2} + a_t - 0.25a_{t-1}, \qquad a_t \sim N(0, 1.1^2).$$

The state-space form of the model is

```
> arma21.m = list(ar=c(1.2,-0.35),ma=c(-0.25),sigma=1.1)
> ssf.arma21= GetSsfArma(model=arma21.m)
> ssf.arma21
$mPhi:
      [,1] [,2]
[1,]  1.20    1
[2,] -0.35    0
[3,]  1.00    0
$mOmega:
        [,1]      [,2] [,3]
[1,]  1.2100 -0.302500    0
[2,] -0.3025  0.075625    0
[3,]  0.0000  0.000000    0
$mSigma:
          [,1]       [,2]
[1,]  4.060709 -1.4874057
[2,] -1.487406  0.5730618
[3,]  0.000000  0.0000000
```

As expected, the output shows that

$$\boldsymbol{T} = \begin{bmatrix} 1.2 & 1 \\ -0.35 & 0 \end{bmatrix}, \qquad \boldsymbol{Z} = (1, 0),$$

and `mPhi` and `mOmega` follow the format of Eq. (11.28), and the covariance matrix of $(s_{1t}, s_{2t})'$ is used in `mSigma`, where $s_{1t} = y_t$ and $s_{2t} = -0.35y_{t-1} - 0.25y_{t-2}$. Note that in SsfPack, the MA polynomial of an ARMA model assumes the form $\theta(B) = 1 + \theta_1 B + \ldots + \theta_q B^q$, not the form $\theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q$ commonly used in the literature.

11.3.3 Linear Regression Model

Multiple linear regression models can also be represented in state-space form. Consider the model

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + e_t, \qquad e_t \sim N(0, \sigma_e^2),$$

where $\boldsymbol{x}_t$ is a p-dimensional explanatory variable and $\boldsymbol{\beta}$ is a p-dimensional parameter vector. Let $\boldsymbol{s}_t = \boldsymbol{\beta}$ for all t. Then the model can be written as

$$\begin{bmatrix} \boldsymbol{s}_{t+1} \\ y_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_p \\ \boldsymbol{x}_t' \end{bmatrix} \boldsymbol{s}_t + \begin{bmatrix} \boldsymbol{0}_p \\ e_t \end{bmatrix}. \tag{11.46}$$

Thus, the system matrices are $\boldsymbol{T}_t = \boldsymbol{I}_p$, $\boldsymbol{Z}_t = \boldsymbol{x}_t'$, $\boldsymbol{d}_t = \boldsymbol{0}$, $c_t = 0$, $\boldsymbol{Q}_t = \boldsymbol{0}$, and $\boldsymbol{H}_t = \sigma_e^2$. Since the state vector is fixed, a diffuse initialization should be used.

One can extend the regression model so that $\boldsymbol{\beta}_t$ is random, say,

$$\boldsymbol{\beta}_{t+1} = \boldsymbol{\beta}_t + \boldsymbol{R}_t \eta_t, \qquad \eta_t \sim N(0, 1),$$

and $\boldsymbol{R}_t = (\sigma_1, \ldots, \sigma_p)'$ with $\sigma_i \geq 0$. If $\sigma_i = 0$, then β_i is time invariant.

SsfPack Command

In SsfPack, the command `GetSsfReg` creates a state-space form for the multiple linear regression model. The command has an input argument that contains the data matrix of explanatory variables. To illustrate, consider the simple market model

$$r_t = \beta_0 + \beta_1 r_{M,t} + e_t, \qquad t = 1, \ldots, 168,$$

where r_t is the return of an asset and $r_{M,t}$ is the market return, for example, the S&P 500 composite index return. The state-space form can be obtained as

```
> ssf.reg=GetSsfReg(cbind(1,sp)) % 'sp' is market return.
> ssf.reg
$mPhi:
     [,1] [,2]
[1,]    1    0
[2,]    0    1
[3,]    0    0
$mOmega:
     [,1] [,2] [,3]
[1,]    0    0    0
[2,]    0    0    0
[3,]    0    0    1
$mSigma:
     [,1] [,2]
[1,]   -1    0
[2,]    0   -1
[3,]    0    0
$mJPhi:
     [,1] [,2]
[1,]   -1   -1
[2,]   -1   -1
[3,]    1    2
$mX:
```

```
numeric matrix: 168 rows, 2 columns.
              sp
 [1,] 1 -0.075187
...
[168,] 1  0.05002
```

11.3.4 Linear Regression Models with ARMA Errors

Consider the regression model with ARMA(p, q) errors:

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + z_t, \qquad \phi(B)z_t = \theta(B)a_t, \tag{11.47}$$

where $a_t \sim N(0, \sigma_a^2)$ and $\boldsymbol{x}_t$ is a k-dimensional vector of explanatory variables. A special case of this model is the nonzero mean ARMA(p, q) model in which $\boldsymbol{x}_t = 1$ for all t and $\boldsymbol{\beta}$ becomes a scalar parameter. Let $\boldsymbol{s}_t$ be a state vector for the z_t series, for example, that defined in Eq. (11.39). We can define a state vector $\boldsymbol{s}_t^*$ for y_t as

$$\boldsymbol{s}_t^* = \begin{bmatrix} \boldsymbol{s}_t \\ \boldsymbol{\beta}_t \end{bmatrix}, \tag{11.48}$$

where $\boldsymbol{\beta}_t = \boldsymbol{\beta}$ for all t. Then, a state-space form for y_t is

$$\boldsymbol{s}_{t+1}^* = \boldsymbol{T}^*\boldsymbol{s}_t^* + \boldsymbol{R}^*\eta_t, \tag{11.49}$$

$$y_t = \boldsymbol{Z}_t^*\boldsymbol{s}_t^*, \tag{11.50}$$

where $\boldsymbol{Z}_t^* = (1, 0, \ldots, 0, \boldsymbol{x}_t')_{1\times(m+k)}$, $m = \max(p, q+1)$, and

$$\boldsymbol{T}^* = \begin{bmatrix} \boldsymbol{T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_k \end{bmatrix}, \qquad \boldsymbol{R}^* = \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{0} \end{bmatrix},$$

where $\boldsymbol{T}$ and $\boldsymbol{R}$ are defined in Eq. (11.39). In a compact form, we have the state-space model

$$\begin{bmatrix} \boldsymbol{s}_{t+1}^* \\ y_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{T}^* \\ \boldsymbol{Z}_t^* \end{bmatrix} \boldsymbol{s}_t^* + \begin{bmatrix} \boldsymbol{R}^*\eta_t \\ 0 \end{bmatrix}.$$

SsfPack Command

SsdPack uses the command `GetSsfRegArma` to construct a state-space form for linear regression models with `ARMA` errors. The arguments of the command can be found using the command `args(GetSsfRegArma)`. They consist of a data matrix for the explanatory variables and ARMA model specification. To illustrate, consider the model

$$y_t = \beta_0 + \beta_1 x_t + z_t, \qquad t = 1, \cdots, 168,$$
$$z_t = 1.2z_{t-1} - 0.35z_{t-2} + a_t - 0.25a_{t-1}, \qquad a_t \sim N(0, \sigma_a^2).$$

We use the notation X to denote the $T \times 2$ matrix of regressors $(1, x_t)$. A state-space form for the prior model can be obtained as

```
> ssf.reg.arma21=GetSsfRegArma(X,ar=c(1.2,-0.35),
+ ma=c(-0.25))
> ssf.reg.arma21
$mPhi:
      [,1] [,2] [,3] [,4]
[1,]  1.20    1    0    0
[2,] -0.35    0    0    0
[3,]  0.00    0    1    0
[4,]  0.00    0    0    1
[5,]  1.00    0    0    0
$mOmega:
      [,1]    [,2] [,3] [,4] [,5]
[1,]  1.00 -0.2500    0    0    0
[2,] -0.25  0.0625    0    0    0
[3,]  0.00  0.0000    0    0    0
[4,]  0.00  0.0000    0    0    0
[5,]  0.00  0.0000    0    0    0
$mSigma:
         [,1]      [,2] [,3] [,4]
[1,]  3.35595 -1.229260    0    0
[2,] -1.22926  0.473604    0    0
[3,]  0.00000  0.000000   -1    0
[4,]  0.00000  0.000000    0   -1
[5,]  0.00000  0.000000    0    0
$mJPhi:
     [,1] [,2] [,3] [,4]
[1,]   -1   -1   -1   -1
[2,]   -1   -1   -1   -1
[3,]   -1   -1   -1   -1
[4,]   -1   -1   -1   -1
[5,]   -1   -1    1    2
$mX:
numeric matrix: 168 rows, 2 columns.
           xt
  [1,] 1 0.4993
...
[168,] 1 0.7561
```

11.3.5 Scalar Unobserved Component Model

The basic univariate unobserved component model, or the *structural time series model* (STSM), assumes the form

$$y_t = \mu_t + \gamma_t + \overline{\omega}_t + e_t, \tag{11.51}$$

where μ_t, γ_t, and $\overline{\omega}_t$ represent the unobserved *trend, seasonal*, and *cycle* components, respectively, and e_t is the unobserved *irregular* component. In the literature, a nonstationary (possibly double-unit-root) model is commonly used for the trend component:

$$\begin{aligned} \mu_{t+1} &= \mu_t + \beta_t + \eta_t, \qquad \eta_t \sim N(0, \sigma_\eta^2), \\ \beta_t &= \beta_{t-1} + \varsigma_t, \quad \varsigma_t \sim N(0, \sigma_\varsigma^2), \end{aligned} \tag{11.52}$$

where $\mu_1 \sim N(0, \xi)$ and $\beta_1 \sim N(0, \xi)$ with ξ a large real number, for example, $\xi = 10^8$. See, for instance, Kitagawa and Gersch (1996). If $\sigma_\varsigma = 0$, then μ_t follows a random walk with drift β_1. If $\sigma_\varsigma = \sigma_\eta = 0$, then μ_t represents a linear deterministic trend.

The seasonal component γ_t assumes the form

$$(1 + B + \cdots + B^{s-1})\gamma_t = \omega_t, \qquad \omega_t \sim N(0, \sigma_\omega^2), \tag{11.53}$$

where s is the number of seasons in a year, that is, the period of the seasonality. If $\sigma_\omega = 0$, then the seasonal pattern is deterministic. The cycle component is postulated as

$$\begin{bmatrix} \overline{\omega}_{t+1} \\ \overline{\omega}^*_{t+1} \end{bmatrix} = \delta \begin{bmatrix} \cos(\lambda_c) & \sin(\lambda_c) \\ -\sin(\lambda_c) & \cos(\lambda_c) \end{bmatrix} \begin{bmatrix} \overline{\omega}_t \\ \overline{\omega}^*_t \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \varepsilon^*_t \end{bmatrix}, \tag{11.54}$$

where

$$\begin{bmatrix} \varepsilon_t \\ \varepsilon^*_t \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma_\varepsilon^2 (1 - \delta^2) \boldsymbol{I}_2 \right),$$

$\overline{\omega}_0 \sim N(0, \sigma_\varepsilon^2)$, $\overline{\omega}^*_0 \sim N(0, \sigma_\varepsilon^2)$, and $\text{Cov}(\overline{\omega}_0, \overline{\omega}^*_0) = 0$, $\delta \in (0, 1]$ is called a *damping* factor, and the frequency of the cycle is $\lambda_c = 2\pi/q$ with q being the period. If $\delta = 1$, then the cycle becomes a deterministic sine–cosine wave.

SsfPack/S-Plus Command

The command `GetSsfStsm` constructs a state-space form for the structural time series model. It allows for 10 cycle components; see the output of the command `args(GetSsfStsm)`. Table 11.4 provides a summary of the arguments and their corresponding symbols of the model. To illustrate, consider the local trend model in Eqs. (11.1) and (11.2) with $\sigma_e = 0.4$ and $\sigma_\eta = 0.2$. This is a special case of the scalar unobserved component model. One can obtain a state-space form as

TABLE 11.4 Arguments of Command GetSsfStsm in SsfPack/S-Plus

Argument	STSM parameter
irregular	σ_e
level	σ_η
slope	σ_ς
seasonalDummy	σ_ω, s
seasonalTrig	σ_ω, s
SeasonalHS	σ_ω, s
Cycle0	σ_ε, λ_c, δ
$\vdots$	$\vdots$
Cycle9	σ_ε, λ_c, δ

```
> ssf.stsm=GetSsfStsm(irregular=0.4,level=0.2)
> ssf.stsm
$mPhi:
     [,1]
[1,]    1
[2,]    1
$mOmega:
     [,1] [,2]
[1,] 0.04 0.00
[2,] 0.00 0.16
$mSigma:
     [,1]
[1,]   -1
[2,]    0
```

11.4 KALMAN FILTER AND SMOOTHING

In this section, we study the Kalman filter and various smoothing methods for the general state-space model in Eqs. (11.26) and (11.27). The derivation follows closely the steps taken in Section 11.1. For readers interested in applications, this section can be skipped at the first read. A good reference for this section is Durbin and Koopman (2001, Chapter 4).

11.4.1 Kalman Filter

Recall that the aim of the Kalman filter is to obtain recursively the conditional distribution of $\boldsymbol{s}_{t+1}$ given the data $F_t = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t\}$ and the model. Since the conditional distribution involved is normal, it suffices to study the conditional mean and covariance matrix. Let $\boldsymbol{s}_{j|i}$ and $\boldsymbol{\Sigma}_{j|i}$ be the conditional mean and covariance matrix of $\boldsymbol{s}_j$ given F_i, that is, $\boldsymbol{s}_j|F_i \sim N(\boldsymbol{s}_{j|i}, \boldsymbol{\Sigma}_{j|i})$. From Eq. (11.26),

$$\boldsymbol{s}_{t+1|t} = E(\boldsymbol{d}_t + \boldsymbol{T}_t\boldsymbol{s}_t + \boldsymbol{R}_t\boldsymbol{\eta}_t|F_t) = \boldsymbol{d}_t + \boldsymbol{T}_t\boldsymbol{s}_{t|t}, \tag{11.55}$$

$$\boldsymbol{\Sigma}_{t+1|t} = \text{Var}(\boldsymbol{T}_t\boldsymbol{s}_t + \boldsymbol{R}_t\boldsymbol{\eta}_t|F_t) = \boldsymbol{T}_t\boldsymbol{\Sigma}_{t|t}\boldsymbol{T}_t' + \boldsymbol{R}_t\boldsymbol{Q}_t\boldsymbol{R}_t'. \tag{11.56}$$

Similarly to that of Section 11.1, let $\boldsymbol{y}_{t|t-1}$ be the conditional mean of y_t given F_{t-1}. From Eq. (11.27),

$$\boldsymbol{y}_{t|t-1} = \boldsymbol{c}_t + \boldsymbol{Z}_t\boldsymbol{s}_{t|t-1}.$$

Let

$$\boldsymbol{v}_t = \boldsymbol{y}_t - \boldsymbol{y}_{t|t-1} = \boldsymbol{y}_t - (\boldsymbol{c}_t + \boldsymbol{Z}_t\boldsymbol{s}_{t|t-1}) = \boldsymbol{Z}_t(\boldsymbol{s}_t - \boldsymbol{s}_{t|t-1}) + \boldsymbol{e}_t, \tag{11.57}$$

be the 1-step-ahead forecast error of $\boldsymbol{y}_t$ given F_{t-1}. It is easy to see that (a) $E(\boldsymbol{v}_t|F_{t-1}) = \boldsymbol{0}$; (b) $\boldsymbol{v}_t$ is independent of F_{t-1}, that is, $\text{Cov}(\boldsymbol{v}_t, \boldsymbol{y}_j) = \boldsymbol{0}$ for $1 \le j < t$; and (c) $\{\boldsymbol{v}_t\}$ is a sequence of independent normal random vectors. Also, let $\boldsymbol{V}_t = \text{Var}(\boldsymbol{v}_t|F_{t-1}) = \text{Var}(\boldsymbol{v}_t)$ be the covariance matrix of the 1-step-ahead forecast error. From Eq. (11.57), we have

$$\boldsymbol{V}_t = \text{Var}[\boldsymbol{Z}_t(\boldsymbol{s}_t - \boldsymbol{s}_{t|t-1}) + \boldsymbol{e}_t] = \boldsymbol{Z}_t\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{Z}_t' + \boldsymbol{H}_t. \tag{11.58}$$

Since $F_t = \{F_{t-1}, \boldsymbol{y}_t\} = \{F_{t-1}, \boldsymbol{v}_t\}$, we can apply Theorem 11.1 to obtain

$$\begin{aligned}\boldsymbol{s}_{t|t} &= E(\boldsymbol{s}_t|F_t) = E(\boldsymbol{s}_t|F_{t-1}, \boldsymbol{v}_t)\\ &= E(\boldsymbol{s}_t|F_{t-1}) + \text{Cov}(\boldsymbol{s}_t, \boldsymbol{v}_t)[\text{Var}(\boldsymbol{v}_t)]^{-1}\boldsymbol{v}_t\\ &= \boldsymbol{s}_{t|t-1} + \boldsymbol{C}_t\boldsymbol{V}_t^{-1}\boldsymbol{v}_t,\end{aligned} \tag{11.59}$$

where $\boldsymbol{C}_t = \text{Cov}(\boldsymbol{s}_t, \boldsymbol{v}_t|F_{t-1})$ given by

$$\begin{aligned}\boldsymbol{C}_t &= \text{Cov}(\boldsymbol{s}_t, \boldsymbol{v}_t|F_{t-1}) = \text{Cov}[\boldsymbol{s}_t, \boldsymbol{Z}_t(\boldsymbol{s}_t - \boldsymbol{s}_{t|t-1}) + \boldsymbol{e}_t|F_{t-1}]\\ &= \text{Cov}[\boldsymbol{s}_t, \boldsymbol{Z}_t(\boldsymbol{s}_t - \boldsymbol{s}_{t|t-1})|F_{t-1}] = \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{Z}_t'.\end{aligned}$$

Here we assume that $\boldsymbol{V}_t$ is invertible because $\boldsymbol{H}_t$ is. Using Eqs. (11.55) and (11.59), we obtain

$$\boldsymbol{s}_{t+1|t} = \boldsymbol{d}_t + \boldsymbol{T}_t\boldsymbol{s}_{t|t-1} + \boldsymbol{T}_t\boldsymbol{C}_t\boldsymbol{V}_t^{-1}\boldsymbol{v}_t = \boldsymbol{d}_t + \boldsymbol{T}_t\boldsymbol{s}_{t|t-1} + \boldsymbol{K}_t\boldsymbol{v}_t, \tag{11.60}$$

where

$$\boldsymbol{K}_t = \boldsymbol{T}_t\boldsymbol{C}_t\boldsymbol{V}_t^{-1} = \boldsymbol{T}_t\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{Z}_t'\boldsymbol{V}_t^{-1}, \tag{11.61}$$

which is the *Kalman gain* at time t. Applying Theorem 11.1(2), we have

$$\begin{aligned}
\boldsymbol{\Sigma}_{t|t} &= \mathrm{Var}(\boldsymbol{s}_t|F_{t-1}, \boldsymbol{v}_t) \\
&= \mathrm{Var}(\boldsymbol{s}_t|F_{t-1}) - \mathrm{Cov}(\boldsymbol{s}_t, \boldsymbol{v}_t)[\mathrm{Var}(\boldsymbol{v}_t)]^{-1}\mathrm{Cov}(\boldsymbol{s}_t, \boldsymbol{v}_t)' \\
&= \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{C}_t\boldsymbol{V}_t^{-1}\boldsymbol{C}_t' \\
&= \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{Z}_t'\boldsymbol{V}_t^{-1}\boldsymbol{Z}_t\boldsymbol{\Sigma}_{t|t-1}.
\end{aligned} \tag{11.62}$$

Plugging Eq. (11.62) into Eq. (11.56) and using Eq. (11.61), we obtain

$$\boldsymbol{\Sigma}_{t+1|t} = \boldsymbol{T}_t\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{L}_t' + \boldsymbol{R}_t\boldsymbol{Q}_t\boldsymbol{R}_t', \tag{11.63}$$

where

$$\boldsymbol{L}_t = \boldsymbol{T}_t - \boldsymbol{K}_t\boldsymbol{Z}_t.$$

Putting the prior equations together, we obtain the celebrated Kalman filter for the state-space model in Eqs. (11.26) and (11.27). Given the starting values $\boldsymbol{s}_{1|0}$ and $\boldsymbol{\Sigma}_{1|0}$, the Kalman filter algorithm is

$$\begin{aligned}
\boldsymbol{v}_t &= \boldsymbol{y}_t - \boldsymbol{c}_t - \boldsymbol{Z}_t\boldsymbol{s}_{t|t-1}, \\
\boldsymbol{V}_t &= \boldsymbol{Z}_t\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{Z}_t' + \boldsymbol{H}_t, \\
\boldsymbol{K}_t &= \boldsymbol{T}_t\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{Z}_t'\boldsymbol{V}_t^{-1}, \\
\boldsymbol{L}_t &= \boldsymbol{T}_t - \boldsymbol{K}_t\boldsymbol{Z}_t, \\
\boldsymbol{s}_{t+1|t} &= \boldsymbol{d}_t + \boldsymbol{T}_t\boldsymbol{s}_{t|t-1} + \boldsymbol{K}_t\boldsymbol{v}_t, \\
\boldsymbol{\Sigma}_{t+1|t} &= \boldsymbol{T}_t\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{L}_t' + \boldsymbol{R}_t\boldsymbol{Q}_t\boldsymbol{R}_t', \qquad t = 1, \ldots, T.
\end{aligned} \tag{11.64}$$

If the filtered quantities $\boldsymbol{s}_{t|t}$ and $\boldsymbol{\Sigma}_{t|t}$ are also of interest, then we modify the filter to include the contemporaneous filtering equations in Eqs. (11.59) and (11.62). The resulting algorithm is

$$\begin{aligned}
\boldsymbol{v}_t &= \boldsymbol{y}_t - \boldsymbol{c}_t - \boldsymbol{Z}_t\boldsymbol{s}_{t|t-1}, \\
\boldsymbol{C}_t &= \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{Z}_t', \\
\boldsymbol{V}_t &= \boldsymbol{Z}_t\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{Z}_t' + \boldsymbol{H}_t = \boldsymbol{Z}_t\boldsymbol{C}_t + \boldsymbol{H}_t, \\
\boldsymbol{s}_{t|t} &= \boldsymbol{s}_{t|t-1} + \boldsymbol{C}_t\boldsymbol{V}_t^{-1}\boldsymbol{v}_t, \\
\boldsymbol{\Sigma}_{t|t} &= \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{C}_t\boldsymbol{V}_t^{-1}\boldsymbol{C}_t', \\
\boldsymbol{s}_{t+1|t} &= \boldsymbol{d}_t + \boldsymbol{T}_t\boldsymbol{s}_{t|t}, \\
\boldsymbol{\Sigma}_{t+1|t} &= \boldsymbol{T}_t\boldsymbol{\Sigma}_{t|t}\boldsymbol{T}_t' + \boldsymbol{R}_t\boldsymbol{Q}_t\boldsymbol{R}_t'.
\end{aligned}$$

Steady State

If the state-space model is time invariant, that is, all system matrices are time invariant, then the matrices $\boldsymbol{\Sigma}_{t|t-1}$ converge to a constant matrix $\boldsymbol{\Sigma}_*$, which is a solution of the matrix equation

$$\boldsymbol{\Sigma}_* = \boldsymbol{T}\boldsymbol{\Sigma}_*\boldsymbol{T}' - \boldsymbol{T}\boldsymbol{\Sigma}_*\boldsymbol{Z}\boldsymbol{V}^{-1}\boldsymbol{Z}\boldsymbol{\Sigma}_*\boldsymbol{T}' + \boldsymbol{R}\boldsymbol{Q}\boldsymbol{R}',$$

where $\boldsymbol{V} = \boldsymbol{Z}\boldsymbol{\Sigma}_*\boldsymbol{Z}' + \boldsymbol{H}$. The solution that is reached after convergence to $\boldsymbol{\Sigma}_*$ is referred to as the *steady-state solution* of the Kalman filter. Once the steady state is reached, $\boldsymbol{V}_t$, $\boldsymbol{K}_t$, and $\boldsymbol{\Sigma}_{t+1|t}$ are all constant. This can lead to considerable saving in computing time.

11.4.2 State Estimation Error and Forecast Error

Define the state prediction error as

$$\boldsymbol{x}_t = \boldsymbol{s}_t - \boldsymbol{s}_{t|t-1}.$$

From the definition, the covariance matrix of $\boldsymbol{x}_t$ is $\mathrm{Var}(\boldsymbol{x}_t|F_{t-1}) = \mathrm{Var}(\boldsymbol{s}_t|F_{t-1}) = \boldsymbol{\Sigma}_{t|t-1}$. Following Section 11.1, we investigate properties of $\boldsymbol{x}_t$. First, from Eq. (11.57),

$$\boldsymbol{v}_t = \boldsymbol{Z}_t(\boldsymbol{s}_t - \boldsymbol{s}_{t|t-1}) + \boldsymbol{e}_t = \boldsymbol{Z}_t\boldsymbol{x}_t + \boldsymbol{e}_t.$$

Second, from Eqs. (11.64) and (11.26), and the prior equation, we have

$$\begin{aligned}
\boldsymbol{x}_{t+1} &= \boldsymbol{s}_{t+1} - \boldsymbol{s}_{t+1|t} \\
&= \boldsymbol{T}_t(\boldsymbol{s}_t - \boldsymbol{s}_{t|t-1}) + \boldsymbol{R}_t\boldsymbol{\eta}_t - \boldsymbol{K}_t\boldsymbol{v}_t \\
&= \boldsymbol{T}_t\boldsymbol{x}_t + \boldsymbol{R}_t\boldsymbol{\eta}_t - \boldsymbol{K}_t(\boldsymbol{Z}_t\boldsymbol{x}_t + \boldsymbol{e}_t) \\
&= \boldsymbol{L}_t\boldsymbol{x}_t + \boldsymbol{R}_t\boldsymbol{\eta}_t - \boldsymbol{K}_t\boldsymbol{e}_t,
\end{aligned}$$

where, as before, $\boldsymbol{L}_t = \boldsymbol{T}_t - \boldsymbol{K}_t\boldsymbol{Z}_t$. Consequently, we obtain a state-space form for $\boldsymbol{v}_t$ as

$$\boldsymbol{v}_t = \boldsymbol{Z}_t\boldsymbol{x}_t + \boldsymbol{e}_t, \qquad \boldsymbol{x}_{t+1} = \boldsymbol{L}_t\boldsymbol{x}_t + \boldsymbol{R}_t\boldsymbol{\eta}_t - \boldsymbol{K}_t\boldsymbol{e}_t, \tag{11.65}$$

with $\boldsymbol{x}_1 = \boldsymbol{s}_1 - \boldsymbol{s}_{1|0}$ for $t = 1, \ldots, T$.

Finally, similar to the local-trend model in Section 11.1, we can show that the 1-step-ahead forecast errors $\{\boldsymbol{v}_t\}$ are independent of each other and $\{\boldsymbol{v}_t, \ldots, \boldsymbol{v}_T\}$ is independent of F_{t-1}.

11.4.3 State Smoothing

State smoothing focuses on the conditional distribution of $\boldsymbol{s}_t$ given F_T. Notice that (a) F_{t-1} and $\{\boldsymbol{v}_t, \ldots, \boldsymbol{v}_T\}$ are independent and (b) $\boldsymbol{v}_t$ are serially independent. We can apply Theorem 11.1 to the joint distribution of $\boldsymbol{s}_t$ and $\{\boldsymbol{v}_t, \ldots, \boldsymbol{v}_T\}$ given F_{t-1} and obtain

$$\begin{aligned}
\boldsymbol{s}_{t|T} &= E(\boldsymbol{s}_t|F_T) = E(\boldsymbol{s}_t|F_{t-1}, \boldsymbol{v}_t, \ldots, \boldsymbol{v}_T) \\
&= E(\boldsymbol{s}_t|F_{t-1}) + \sum_{j=t}^{T} \text{Cov}(\boldsymbol{s}_t, \boldsymbol{v}_j)[\text{Var}(\boldsymbol{v}_j)]^{-1}\boldsymbol{v}_j \\
&= \boldsymbol{s}_{t|t-1} + \sum_{j=t}^{T} \text{Cov}(\boldsymbol{s}_t, \boldsymbol{v}_j)\boldsymbol{V}_j^{-1}\boldsymbol{v}_j,
\end{aligned} \tag{11.66}$$

where the covariance matrices are conditional on F_{t-1}. The covariance matrices $\text{Cov}(\boldsymbol{s}_t, \boldsymbol{v}_j)$ for $j = t, \ldots, T$ can be derived as follows. By Eq. (11.65),

$$\begin{aligned}
\text{Cov}(\boldsymbol{s}_t, \boldsymbol{v}_j) &= E(\boldsymbol{s}_t\boldsymbol{v}_j') \\
&= E[\boldsymbol{s}_t(\boldsymbol{Z}_j\boldsymbol{x}_j + \boldsymbol{e}_j)'] = E(\boldsymbol{s}_t\boldsymbol{x}_j')\boldsymbol{Z}_j', \qquad j = t, \ldots, T.
\end{aligned} \tag{11.67}$$

Furthermore,

$$\begin{aligned}
E(\boldsymbol{s}_t\boldsymbol{x}_t') &= E[\boldsymbol{s}_t(\boldsymbol{s}_t - \boldsymbol{s}_{t|t-1})'] = \text{Var}(\boldsymbol{s}_t) = \boldsymbol{\Sigma}_{t|t-1}, \\
E(\boldsymbol{s}_t\boldsymbol{x}_{t+1}') &= E[\boldsymbol{s}_t(\boldsymbol{L}_t\boldsymbol{x}_t + \boldsymbol{R}_t\boldsymbol{\eta}_t - \boldsymbol{K}_t\boldsymbol{e}_t)'] = \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{L}_t', \\
E(\boldsymbol{s}_t\boldsymbol{x}_{t+2}') &= \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{L}_t'\boldsymbol{L}_{t+1}', \\
&\;\;\vdots \\
E(\boldsymbol{s}_t\boldsymbol{x}_T') &= \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{L}_t' \cdots \boldsymbol{L}_{T-1}'.
\end{aligned} \tag{11.68}$$

Plugging the prior two equations into Eq. (11.66), we have

$$\begin{aligned}
\boldsymbol{s}_{T|T} &= \boldsymbol{s}_{T|T-1} + \boldsymbol{\Sigma}_{T|T-1}\boldsymbol{Z}_T'\boldsymbol{V}_T^{-1}\boldsymbol{v}_T, \\
\boldsymbol{s}_{T-1|T} &= \boldsymbol{s}_{T-1|T-2} + \boldsymbol{\Sigma}_{T-1|T-2}\boldsymbol{Z}_{T-1}'\boldsymbol{V}_{T-1}^{-1}\boldsymbol{v}_{T-1} + \boldsymbol{\Sigma}_{T-1|T-2}\boldsymbol{L}_{T-1}'\boldsymbol{Z}_T'\boldsymbol{V}_T^{-1}\boldsymbol{v}_T, \\
\boldsymbol{s}_{t|T} &= \boldsymbol{s}_{t|t-1} + \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{Z}_t'\boldsymbol{V}_t^{-1}\boldsymbol{v}_t + \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{L}_t'\boldsymbol{Z}_{t+1}'\boldsymbol{V}_{t+1}^{-1}\boldsymbol{v}_{t+1} \\
&\quad + \cdots + \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{L}_t'\boldsymbol{L}_{t+1}' \cdots \boldsymbol{L}_{T-1}'\boldsymbol{Z}_T'\boldsymbol{V}_T^{-1}\boldsymbol{v}_T,
\end{aligned}$$

for $t = T-2, T-3, \ldots, 1$, where it is understood that $\boldsymbol{L}_t' \cdots \boldsymbol{L}_{T-1}' = \boldsymbol{I}_m$ when $t = T$. These smoothed state vectors can be expressed as

$$\boldsymbol{s}_{t|T} = \boldsymbol{s}_{t|t-1} + \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{q}_{t-1}, \tag{11.69}$$

where $\boldsymbol{q}_{T-1} = \boldsymbol{Z}_T'\boldsymbol{V}_T^{-1}\boldsymbol{v}_T$, $\boldsymbol{q}_{T-2} = \boldsymbol{Z}_{T-1}'\boldsymbol{V}_{T-1}^{-1}\boldsymbol{v}_{T-1} + \boldsymbol{L}_{T-1}'\boldsymbol{Z}_T'\boldsymbol{V}_T^{-1}\boldsymbol{v}_T$, and

$$\boldsymbol{q}_{t-1} = \boldsymbol{Z}_t'\boldsymbol{V}_t^{-1}\boldsymbol{v}_t + \boldsymbol{L}_t'\boldsymbol{Z}_{t+1}'\boldsymbol{V}_{t+1}^{-1}\boldsymbol{v}_{t+1} + \cdots + \boldsymbol{L}_t'\boldsymbol{L}_{t+1}'\cdots\boldsymbol{L}_{T-1}'\boldsymbol{Z}_T'\boldsymbol{V}_T^{-1}\boldsymbol{v}_T,$$

for $t = T-2, T-3, \ldots, 1$. The quantity $\boldsymbol{q}_{t-1}$ is a weighted sum of the 1-step-ahead forecast errors $\boldsymbol{v}_j$ occurring after time $t-1$. From the definition in the prior equation, $\boldsymbol{q}_t$ can be computed recursively backward as

$$\boldsymbol{q}_{t-1} = \boldsymbol{Z}_t'\boldsymbol{V}_t^{-1}\boldsymbol{v}_t + \boldsymbol{L}_t'\boldsymbol{q}_t, \qquad t = T, \ldots, 1, \tag{11.70}$$

with $\boldsymbol{q}_T = \boldsymbol{0}$. Putting the equations together, we have a backward recursion for the smoothed state vectors as

$$\boldsymbol{q}_{t-1} = \boldsymbol{Z}_t'\boldsymbol{V}_t^{-1}\boldsymbol{v}_t + \boldsymbol{L}_t'\boldsymbol{q}_t, \qquad \boldsymbol{s}_{t|T} = \boldsymbol{s}_{t|t-1} + \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{q}_{t-1}, \qquad t = T, \ldots, 1, \tag{11.71}$$

starting with $\boldsymbol{q}_T = \boldsymbol{0}$, where $\boldsymbol{s}_{t|t-1}$, $\boldsymbol{\Sigma}_{t|t-1}$, $\boldsymbol{L}_t$, and $\boldsymbol{V}_t$ are available from the Kalman filter. This algorithm is referred to as the *fixed interval smoother* in the literature; see de Jong (1989) and the references therein.

Covariance Matrix of Smoothed State Vector

Next, we derive the covariance matrices of the smoothed state vectors. Applying Theorem 11.1(4) to the conditional joint distribution of $\boldsymbol{s}_t$ and $\{\boldsymbol{v}_t, \ldots, \boldsymbol{v}_T\}$ given F_{t-1}, we have

$$\boldsymbol{\Sigma}_{t|T} = \boldsymbol{\Sigma}_{t|t-1} - \sum_{j=t}^{T}\mathrm{Cov}(\boldsymbol{s}_t, \boldsymbol{v}_j)[\mathrm{Var}(\boldsymbol{v}_j)]^{-1}[\mathrm{Cov}(\boldsymbol{s}_t, \boldsymbol{v}_j)]'.$$

Using the covariance matrices in Eqs. (11.67) and (11.68), we further obtain

$$\begin{aligned}\boldsymbol{\Sigma}_{t|T} &= \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{Z}_t'\boldsymbol{V}_t^{-1}\boldsymbol{Z}_t\boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{L}_t'\boldsymbol{Z}_{t+1}'\boldsymbol{V}_{t+1}^{-1}\boldsymbol{Z}_{t+1}\boldsymbol{L}_t\boldsymbol{\Sigma}_{t|t-1}\\ &\quad - \cdots - \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{L}_t'\cdots\boldsymbol{L}_{T-1}'\boldsymbol{Z}_T'\boldsymbol{V}_T^{-1}\boldsymbol{Z}_T\boldsymbol{L}_{T-1}\cdots\boldsymbol{L}_t\boldsymbol{\Sigma}_{t|t-1}\\ &= \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{M}_{t-1}\boldsymbol{\Sigma}_{t|t-1},\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{M}_{t-1} &= \boldsymbol{Z}_t'\boldsymbol{V}_t^{-1}\boldsymbol{Z}_t + \boldsymbol{L}_t'\boldsymbol{Z}_{t+1}'\boldsymbol{V}_{t+1}^{-1}\boldsymbol{Z}_{t+1}\boldsymbol{L}_t\\ &\quad + \cdots + \boldsymbol{L}_t'\cdots\boldsymbol{L}_{T-1}'\boldsymbol{Z}_T'\boldsymbol{V}_T^{-1}\boldsymbol{Z}_T\boldsymbol{L}_{T-1}\cdots\boldsymbol{L}_t.\end{aligned}$$

Again, $\boldsymbol{L}_t'\cdots\boldsymbol{L}_{T-1} = \boldsymbol{I}_m$ when $t = T$. From its definition, the $\boldsymbol{M}_{t-1}$ matrix satisfies

$$\boldsymbol{M}_{t-1} = \boldsymbol{Z}_t'\boldsymbol{V}_t^{-1}\boldsymbol{Z}_t + \boldsymbol{L}_t'\boldsymbol{M}_t\boldsymbol{L}_t, \qquad t = T, \ldots, 1, \tag{11.72}$$

with the starting value $\boldsymbol{M}_T = \boldsymbol{0}$. Collecting the results, we obtain a backward recursion to compute $\boldsymbol{\Sigma}_{t|T}$ as

$$\boldsymbol{M}_{t-1} = \boldsymbol{Z}_t' \boldsymbol{V}_t^{-1} \boldsymbol{Z}_t, + \boldsymbol{L}_t' \boldsymbol{M}_t \boldsymbol{L}_t, \qquad \boldsymbol{\Sigma}_{t|T} = \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{\Sigma}_{t|t-1} \boldsymbol{M}_{t-1} \boldsymbol{\Sigma}_{t|t-1}, \quad (11.73)$$

for $t = T, \ldots, 1$ with $\boldsymbol{M}_T = \boldsymbol{0}$. Note that, like that of the local trend model in Section 11.1, $\boldsymbol{M}_t = \text{Var}(\boldsymbol{q}_t)$.

Combining the two backward recursions of smoothed state vectors, we have

$$\begin{aligned}
\boldsymbol{q}_{t-1} &= \boldsymbol{Z}_t' \boldsymbol{V}_t^{-1} \boldsymbol{v}_t + \boldsymbol{L}_t' \boldsymbol{q}_t, \\
\boldsymbol{s}_{t|T} &= \boldsymbol{s}_{t|t-1} + \boldsymbol{\Sigma}_{t|t-1} \boldsymbol{q}_{t-1}, \\
\boldsymbol{M}_{t-1} &= \boldsymbol{Z}_t' \boldsymbol{V}_t^{-1} \boldsymbol{Z}_t + \boldsymbol{L}_t' \boldsymbol{M}_t \boldsymbol{L}_t, \\
\boldsymbol{\Sigma}_{t|T} &= \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{\Sigma}_{t|t-1} \boldsymbol{M}_{t-1} \boldsymbol{\Sigma}_{t|t-1}, \quad t = T, \ldots, 1,
\end{aligned} \quad (11.74)$$

with $\boldsymbol{q}_T = \boldsymbol{0}$ and $\boldsymbol{M}_T = \boldsymbol{0}$.

Suppose that the state-space model in Eqs. (11.26) and (11.27) is known. Application of the Kalman filter and state smoothing can proceed in two steps. First, the Kalman filter in Eq. (11.64) is used for $t = 1, \ldots, T$ and the quantities $\boldsymbol{v}_t, \boldsymbol{V}_t$, $\boldsymbol{K}_t$, $\boldsymbol{s}_{t|t-1}$, and $\boldsymbol{\Sigma}_{t|t-1}$ are stored. Second, the state smoothing algorithm in Eq. (11.74) is applied for $t = T, T-1, \ldots, 1$ to obtain $\boldsymbol{s}_{t|T}$ and $\boldsymbol{\Sigma}_{t|T}$.

11.4.4 Disturbance Smoothing

Let $\boldsymbol{e}_{t|T} = E(\boldsymbol{e}_t|F_T)$ and $\boldsymbol{\eta}_{t|T} = E(\boldsymbol{\eta}_t|F_T)$ be the smoothed disturbances of the observation and transition equation, respectively. These *smoothed disturbances* are useful in many applications, for example, in model checking. In this section, we study recursive algorithms to compute smoothed disturbances and their covariance matrices. Again, applying Theorem 11.1 to the conditional joint distribution of $\boldsymbol{e}_t$ and $\{\boldsymbol{v}_t, \ldots, \boldsymbol{v}_T\}$ given F_{t-1}, we obtain

$$\boldsymbol{e}_{t|T} = E(\boldsymbol{e}_t|F_{t-1}, \boldsymbol{v}_t, \ldots, \boldsymbol{v}_T) = \sum_{j=t}^{T} E(\boldsymbol{e}_t \boldsymbol{v}_j') \boldsymbol{V}_j^{-1} \boldsymbol{v}_j, \quad (11.75)$$

where $E(\boldsymbol{e}_t|F_{t-1}) = \boldsymbol{0}$ is used. Using Eq. (11.65),

$$E(\boldsymbol{e}_t \boldsymbol{v}_j') = E(\boldsymbol{e}_t \boldsymbol{x}_j') \boldsymbol{Z}_j' + E(\boldsymbol{e}_t \boldsymbol{e}_j').$$

Since $E(\boldsymbol{e}_t \boldsymbol{x}_t') = \boldsymbol{0}$, we have

$$E(\boldsymbol{e}_t \boldsymbol{v}_j') = \begin{cases} \boldsymbol{H}_t, & \text{if } j = t, \\ E(\boldsymbol{e}_t \boldsymbol{x}_j') \boldsymbol{Z}_j', & \text{for } j = t+1, \ldots, T. \end{cases} \quad (11.76)$$

Using Eq. (11.65) repeatedly and the independence between $\{\boldsymbol{e}_t\}$ and $\{\boldsymbol{\eta}_t\}$, we obtain

$$\begin{aligned}
E(\boldsymbol{e}_t\boldsymbol{x}'_{t+1}) &= -\boldsymbol{H}_t\boldsymbol{K}'_t,\\
E(\boldsymbol{e}_t\boldsymbol{x}'_{t+2}) &= -\boldsymbol{H}_t\boldsymbol{K}'_t\boldsymbol{L}'_{t+1},\\
&\vdots\\
E(\boldsymbol{e}_t\boldsymbol{x}'_{T}) &= -\boldsymbol{H}_t\boldsymbol{K}'_t\boldsymbol{L}'_{t+1}\cdots\boldsymbol{L}'_{T-1},
\end{aligned}\tag{11.77}$$

where it is understood that $\boldsymbol{L}'_{t+1}\cdots\boldsymbol{L}'_{T-1}=\boldsymbol{I}_m$ if $t=T-1$. Based on Eqs. (11.76) and (11.77),

$$\begin{aligned}
\boldsymbol{e}_{t|T} &= \boldsymbol{H}_t(\boldsymbol{V}_t^{-1}\boldsymbol{v}_t-\boldsymbol{K}'_t\boldsymbol{Z}'_{t+1}\boldsymbol{V}_{t+1}^{-1}\boldsymbol{v}_{t+1}-\cdots-\boldsymbol{K}'_t\boldsymbol{L}'_{t+1}\cdots\boldsymbol{L}'_{T-1}\boldsymbol{Z}'_T\boldsymbol{V}_T^{-1}\boldsymbol{v}_T)\\
&= \boldsymbol{H}_t(\boldsymbol{V}_t^{-1}\boldsymbol{v}_t-\boldsymbol{K}'_t\boldsymbol{q}_t)\\
&= \boldsymbol{H}_t\boldsymbol{o}_t,\quad t=T,\cdots,1,
\end{aligned}\tag{11.78}$$

where $\boldsymbol{q}_t$ is defined in Eq. (11.69) and $\boldsymbol{o}_t=\boldsymbol{V}_t^{-1}\boldsymbol{v}_t-\boldsymbol{K}'_t\boldsymbol{q}_t$. We refer to $\boldsymbol{o}_t$ as the *smoothing measurement error*.

The smoothed disturbance $\boldsymbol{\eta}_{t|T}$ can be derived analogously, and we have

$$\boldsymbol{\eta}_{t|T}=\sum_{j=t}^{T}E(\boldsymbol{\eta}_t\boldsymbol{v}'_j)\boldsymbol{V}_j^{-1}\boldsymbol{v}_j.\tag{11.79}$$

The state-space form in Eq. (11.69) gives

$$E(\boldsymbol{\eta}_t\boldsymbol{v}'_j)=\begin{cases}\boldsymbol{Q}_t\boldsymbol{R}'_t\boldsymbol{Z}'_{t+1}, & \text{if } j=t+1,\\ E(\boldsymbol{\eta}_t\boldsymbol{x}'_j)\boldsymbol{Z}'_j, & \text{if } j=t+2,\cdots,T,\end{cases}$$

where

$$\begin{aligned}
E(\boldsymbol{\eta}_t\boldsymbol{x}'_{t+2}) &= \boldsymbol{Q}_t\boldsymbol{R}'_t\boldsymbol{L}'_{t+1},\\
E(\boldsymbol{\eta}_t\boldsymbol{x}'_{t+3}) &= \boldsymbol{Q}_t\boldsymbol{R}'_t\boldsymbol{L}'_{t+1}\boldsymbol{L}'_{t+2},\\
&\vdots\\
E(\boldsymbol{\eta}_t\boldsymbol{x}'_{T}) &= \boldsymbol{Q}_t\boldsymbol{R}'_t\boldsymbol{L}'_{t+1}\cdots\boldsymbol{L}'_{T-1},
\end{aligned}$$

for $t=1,\ldots,T$. Consequently, Eq. (11.79) implies

$$\begin{aligned}
\boldsymbol{\eta}_{t|T} &= \boldsymbol{Q}_t\boldsymbol{R}'_t(\boldsymbol{Z}'_{t+1}\boldsymbol{V}_{t+1}^{-1}\boldsymbol{v}_{t+1}+\boldsymbol{L}'_{t+1}\boldsymbol{Z}'_{t+2}\boldsymbol{V}_{t+2}^{-1}\boldsymbol{v}_{t+2}\\
&\qquad+\cdots+\boldsymbol{L}'_{t+1}\cdots\boldsymbol{L}'_{T-1}\boldsymbol{Z}_T\boldsymbol{V}_T^{-1}\boldsymbol{v}_T)\\
&= \boldsymbol{Q}_t\boldsymbol{R}'_t\boldsymbol{q}_t,\qquad t=T,\ldots,1,
\end{aligned}\tag{11.80}$$

where $\boldsymbol{q}_t$ is defined earlier in Eq. (11.70).

Koopman (1993) uses the smoothed disturbance $\boldsymbol{\eta}_{t|T}$ to derive a new recursion for computing $\boldsymbol{s}_{t|T}$. From the transition equation in Eq. (11.26),

$$\boldsymbol{s}_{t+1|T} = \boldsymbol{d}_t + \boldsymbol{T}_t \boldsymbol{s}_{t|T} + \boldsymbol{R}_t \boldsymbol{\eta}_{t|T}.$$

Using Eq. (11.80), we have

$$\boldsymbol{s}_{t+1|T} = \boldsymbol{d}_t + \boldsymbol{T}_t \boldsymbol{s}_{t|T} + \boldsymbol{R}_t \boldsymbol{Q}_t \boldsymbol{R}_t' \boldsymbol{q}_t, \qquad t = 1, \ldots, T, \tag{11.81}$$

where the initial value is $\boldsymbol{s}_{1|T} = \boldsymbol{s}_{1|0} + \boldsymbol{\Sigma}_{1|0} \boldsymbol{q}_0$ with $\boldsymbol{q}_0$ obtained from the recursion in Eq. (11.70).

Covariance Matrices of Smoothed Disturbances

The covariance matrix of the smoothed disturbance can also be obtained using Theorem 11.1. Specifically,

$$\begin{aligned}
\mathrm{Var}(\boldsymbol{e}_t|F_T) &= \mathrm{Var}(\boldsymbol{e}_t|F_{t-1}, \boldsymbol{v}_t, \ldots, \boldsymbol{v}_T) \\
&= \mathrm{Var}(\boldsymbol{e}_t|F_{t-1}) - \sum_{j=t}^{T} \mathrm{Cov}(\boldsymbol{e}_t, \boldsymbol{v}_j) \boldsymbol{V}_j^{-1} [\mathrm{Cov}(\boldsymbol{e}_t, \boldsymbol{v}_j)]'.
\end{aligned}$$

Note that $\mathrm{Cov}(\boldsymbol{e}_t, \boldsymbol{v}_j) = E(\boldsymbol{e}_t \boldsymbol{v}_j')$, which is given in Eq. (11.76). Thus, we have

$$\begin{aligned}
\mathrm{Var}(\boldsymbol{e}_t|F_T) = \boldsymbol{H}_t - \boldsymbol{H}_t (&\boldsymbol{V}_t^{-1} + \boldsymbol{K}_t' \boldsymbol{Z}_{t+1}' \boldsymbol{V}_{t+1}^{-1} \boldsymbol{Z}_{t+1} \boldsymbol{K}_t \\
&+ \boldsymbol{K}_t' \boldsymbol{L}_{t+1}' \boldsymbol{Z}_{t+2}' \boldsymbol{V}_{t+2}^{-1} \boldsymbol{Z}_{t+2} \boldsymbol{L}_{t+1} \boldsymbol{K}_t \\
&+ \cdots + \boldsymbol{K}_t' \boldsymbol{L}_{t+1}' \cdots \boldsymbol{L}_{T-1}' \boldsymbol{Z}_T' \boldsymbol{V}_T^{-1} \boldsymbol{Z}_T \boldsymbol{L}_{T-1} \cdots \boldsymbol{L}_{t+1} \boldsymbol{K}_t) \boldsymbol{H}_t \\
= \boldsymbol{H}_t - \boldsymbol{H}_t (&\boldsymbol{V}_t^{-1} + \boldsymbol{K}_t' \boldsymbol{M}_t \boldsymbol{K}_t) \boldsymbol{H}_t \\
= \boldsymbol{H}_t - \boldsymbol{H}_t &\boldsymbol{N}_t \boldsymbol{H}_t,
\end{aligned}$$

where $\boldsymbol{N}_t = \boldsymbol{V}_t^{-1} + \boldsymbol{K}_t' \boldsymbol{M}_t \boldsymbol{K}_t$, where $\boldsymbol{M}_t$ is given in Eq. (11.72). Similarly,

$$\mathrm{Var}(\boldsymbol{\eta}_t|F_T) = \mathrm{Var}(\boldsymbol{\eta}_t) - \sum_{j=t}^{T} \mathrm{Cov}(\boldsymbol{\eta}_t, \boldsymbol{v}_t) \boldsymbol{V}_t^{-1} [\mathrm{Cov}(\boldsymbol{\eta}_t, \boldsymbol{v}_t)]^{-1},$$

where $\mathrm{Cov}(\boldsymbol{\eta}_t, \boldsymbol{v}_j) = E(\boldsymbol{\eta}_t \boldsymbol{v}_j')$, which is given before when we derived the formula for $\boldsymbol{\eta}_{t|T}$. Consequently,

$$\begin{aligned}
\mathrm{Var}(\boldsymbol{\eta}_t|F_T) &= \boldsymbol{Q}_t - \boldsymbol{Q}_t \boldsymbol{R}_t' (\boldsymbol{Z}_{t+1}' \boldsymbol{V}_{t+1}^{-1} \boldsymbol{Z}_{t+1} + \boldsymbol{L}_{t+1}' \boldsymbol{Z}_{t+2}' \boldsymbol{V}_{t+2}^{-1} \boldsymbol{Z}_{t+2} \boldsymbol{L}_{t+1} \\
&\quad + \cdots + \boldsymbol{L}_{t+1}' \cdots \boldsymbol{L}_{T-1}' \boldsymbol{Z}_T' \boldsymbol{V}_T^{-1} \boldsymbol{Z}_T \boldsymbol{L}_{T-1} \cdots \boldsymbol{L}_{t+1}) \boldsymbol{R}_t \boldsymbol{Q}_t \\
&= \boldsymbol{Q}_t - \boldsymbol{Q}_t \boldsymbol{R}_t' \boldsymbol{M}_t \boldsymbol{R}_t \boldsymbol{Q}_t.
\end{aligned}$$

In summary, the disturbance smoothing algorithm is as follows:

$$\begin{aligned}
\boldsymbol{e}_{t|T} &= \boldsymbol{H}_t(\boldsymbol{V}_t^{-1}\boldsymbol{v}_t - \boldsymbol{K}_t'\boldsymbol{q}_t), \\
\boldsymbol{\eta}_{t|T} &= \boldsymbol{Q}_t\boldsymbol{R}_t'\boldsymbol{q}_t, \\
\boldsymbol{q}_{t-1} &= \boldsymbol{Z}_t'\boldsymbol{V}_t^{-1}\boldsymbol{v}_t + \boldsymbol{L}_t'\boldsymbol{q}_t, \qquad\qquad (11.82)\\
\mathrm{Var}(\boldsymbol{e}_t|F_T) &= \boldsymbol{H}_t - \boldsymbol{H}_t(\boldsymbol{V}_t^{-1} + \boldsymbol{K}_t'\boldsymbol{M}_t\boldsymbol{K}_t)\boldsymbol{H}_t, \\
\mathrm{Var}(\boldsymbol{\eta}_t|F_T) &= \boldsymbol{Q}_t - \boldsymbol{Q}_t\boldsymbol{R}_t'\boldsymbol{M}_t\boldsymbol{R}_t\boldsymbol{Q}_t, \\
\boldsymbol{M}_{t-1} &= \boldsymbol{Z}_t'\boldsymbol{V}_t^{-1}\boldsymbol{Z}_t + \boldsymbol{L}_t'\boldsymbol{M}_t\boldsymbol{L}_t, \qquad t = T, \ldots, 1,
\end{aligned}$$

where $\boldsymbol{q}_T = \boldsymbol{0}$ and $\boldsymbol{M}_T = \boldsymbol{0}$.

11.5 MISSING VALUES

For the general state-space model in Eqs. (11.26) and (11.27), we consider two cases of missing values. First, suppose that similar to the local trend model in Section 11.1 the observations $\boldsymbol{y}_t$ at $t = \ell + 1, \ldots, \ell + h$ are missing. In this case, there is no new information available at these time points and we set

$$\boldsymbol{v}_t = \boldsymbol{0}, \qquad \boldsymbol{K}_t = \boldsymbol{0}, \qquad \text{for} \quad t = \ell + 1, \ldots, \ell + h.$$

The Kalman filter in Eq. (11.64) can then proceed as usual. That is,

$$\boldsymbol{s}_{t+1|t} = \boldsymbol{d}_t + \boldsymbol{T}_t\boldsymbol{s}_{t|t-1}, \qquad \boldsymbol{\Sigma}_{t+1|t} = \boldsymbol{T}_t\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{T}_t' + \boldsymbol{R}_t\boldsymbol{Q}_t\boldsymbol{R}_t',$$

for $t = \ell + 1, \ldots, \ell + h$. Similarly, the smoothed state vectors can be computed as usual via Eq. (11.74) with

$$\boldsymbol{q}_{t-1} = \boldsymbol{T}_t'\boldsymbol{q}_t, \qquad \boldsymbol{M}_{t-1} = \boldsymbol{T}_t'\boldsymbol{M}_t\boldsymbol{T}_t,$$

for $t = \ell + 1, \ldots, \ell + h$.

In the second case, some components of $\boldsymbol{y}_t$ are missing. Let $\boldsymbol{y}_t^* = \boldsymbol{J}\boldsymbol{y}_t$ be the vector of observed data at time t, where $\boldsymbol{J}$ is an indicator matrix identifying the observed data. More specifically, rows of $\boldsymbol{J}$ are a subset of the rows of the $k \times k$ identity matrix. In this case, the observation equation (11.27) of the model can be transformed as

$$\boldsymbol{y}_t^* = \boldsymbol{c}_t^* + \boldsymbol{Z}_t^*\boldsymbol{s}_t + \boldsymbol{e}_t^*,$$

where $\boldsymbol{c}_t^* = \boldsymbol{J}\boldsymbol{c}_t$, $\boldsymbol{Z}_t^* = \boldsymbol{J}\boldsymbol{Z}_t$, and $\boldsymbol{e}_t^* = \boldsymbol{J}\boldsymbol{e}_t$ with covariance matrix $\mathrm{Var}(\boldsymbol{e}_t^*) = \boldsymbol{H}_t^* = \boldsymbol{J}\boldsymbol{H}_t\boldsymbol{J}'$. The Kalman filter and state-smoothing recursion continue to apply except that the modified observation equation is used at time t. Consequently, the ease in handling missing values is a nice feature of the state-space model.

11.6 FORECASTING

Suppose that the forecast origin is t and we are interested in predicting $\boldsymbol{y}_{t+j}$ for $j = 1, \ldots, h$, where $h > 0$. Also, we adopt the minimum mean-squared error forecasts. Similar to the ARMA models, the j-step-ahead forecast $\boldsymbol{y}_t(j)$ turns out to be the expected value of $\boldsymbol{y}_{t+j}$ given F_t and the model. That is, $\boldsymbol{y}_t(j) = E(\boldsymbol{y}_{t+j}|F_t)$. In what follows, we show that these forecasts and the covariance matrices of the associated forecast errors can be obtained via the Kalman filter in Eq. (11.64) by treating $\{\boldsymbol{y}_{t+1}, \ldots, \boldsymbol{y}_{t+h}\}$ as missing values, that is, the first case in Section 11.5.

Consider the 1-step-ahead forecast. From Eq. (11.27),

$$\boldsymbol{y}_t(1) = E(\boldsymbol{y}_{t+1}|F_t) = \boldsymbol{c}_{t+1} + \boldsymbol{Z}_{t+1}\boldsymbol{s}_{t+1|t},$$

where $\boldsymbol{s}_{t+1|t}$ is available via the Kalman filter at the forecast origin t. The associated forecast error is

$$\boldsymbol{e}_t(1) = \boldsymbol{y}_{t+1} - \boldsymbol{y}_t(1) = \boldsymbol{Z}_{t+1}(\boldsymbol{s}_{t+1} - \boldsymbol{s}_{t+1|t}) + \boldsymbol{e}_{t+1}.$$

Therefore, the covariance matrix of the 1-step-ahead forecast error is

$$\text{Var}[\boldsymbol{e}_t(1)] = \boldsymbol{Z}_{t+1}\boldsymbol{\Sigma}_{t+1|t}\boldsymbol{Z}'_{t+1} + \boldsymbol{H}_{t+1}.$$

This is precisely the covariance matrix $\boldsymbol{V}_{t+1}$ of the Kalman filter in Eq. (11.64). Thus, we have showed the case for $h = 1$.

Now, for $h > 1$, we consider 1-step- to h-step-ahead forecasts sequentially. From Eq. (11.27), the j-step-ahead forecast is

$$\boldsymbol{y}_t(j) = \boldsymbol{c}_{t+j} + \boldsymbol{Z}_{t+j}\boldsymbol{s}_{t+j|t}, \tag{11.83}$$

and the associated forecast error is

$$\boldsymbol{e}_t(j) = \boldsymbol{Z}_{t+j}(\boldsymbol{s}_{t+j} - \boldsymbol{s}_{t+j|t}) + \boldsymbol{e}_{t+j}.$$

Recall that $\boldsymbol{s}_{t+j|t}$ and $\boldsymbol{\Sigma}_{t+j|t}$ are, respectively, the conditional mean and covariance matrix of $\boldsymbol{s}_{t+j}$ given F_t. The prior equation says that

$$\text{Var}[\boldsymbol{e}_t(j)] = \boldsymbol{Z}_{t+j}\boldsymbol{\Sigma}_{t+j|t}\boldsymbol{Z}'_{t+j} + \boldsymbol{H}_{t+j}. \tag{11.84}$$

Furthermore, from Eq. (11.26),

$$\boldsymbol{s}_{t+j+1|t} = \boldsymbol{d}_{t+j} + \boldsymbol{T}_{t+j}\boldsymbol{s}_{t+j|t},$$

which in turn implies that

$$\boldsymbol{s}_{t+j+1} - \boldsymbol{s}_{t+j+1|t} = \boldsymbol{T}_{t+j}(\boldsymbol{s}_{t+j} - \boldsymbol{s}_{t+j|t}) + \boldsymbol{R}_{t+j}\boldsymbol{\eta}_{t+j}.$$

Consequently,

$$\boldsymbol{\Sigma}_{t+j+1|t} = \boldsymbol{T}_{t+j}\boldsymbol{\Sigma}_{t+j|t}\boldsymbol{T}'_{t+j} + \boldsymbol{R}_{t+j}\boldsymbol{Q}_{t+j}\boldsymbol{R}'_{t+j}. \tag{11.85}$$

Note that $\text{Var}[\boldsymbol{e}_t(j)] = \boldsymbol{V}_{t+j}$ and Eqs. (11.83) and (11.85) are the recursion of the Kalman filter in Eq. (11.64) for $t+j$ with $j = 1, \ldots, h$ when $\boldsymbol{v}_{t+j} = \boldsymbol{0}$ and $\boldsymbol{K}_{t+j} = \boldsymbol{0}$. Thus, the forecast $\boldsymbol{y}_t(j)$ and the covariance matrix of its forecast error $\boldsymbol{e}_t(j)$ can be obtained via the Kalman filter with missing values.

Finally, the prediction error series $\{\boldsymbol{v}_t\}$ can be used to evaluate the likelihood function for estimation and the standardized prediction errors $\boldsymbol{D}_t^{-1/2}\boldsymbol{v}_t$ can be used for model checking, where $\boldsymbol{D}_t = \text{diag}\{\boldsymbol{V}_t(1,1), \ldots, \boldsymbol{V}_t(k,k)\}$ with $\boldsymbol{V}_t(i,i)$ being the (i,i)th element of $\boldsymbol{V}_t$.

11.7 APPLICATION

In this section, we consider some applications of the state-space model in finance and business. Our objectives are to highlight the applicability of the model and to demonstrate the practical implementation of the analysis in S-Plus with SsfPack.

Example 11.2. Consider the CAPM for the monthly simple excess returns of General Motors (GM) stock from January 1990 to December 2003; see Chapter 9. We use the simple excess returns of the S&P 500 composite index as the market returns. The returns are in percentages. Our illustration starts with a simple market model

$$r_t = \alpha + \beta r_{M,t} + e_t, \qquad e_t \sim N(0, \sigma_e^2), \tag{11.86}$$

for $t = 1, \ldots, 168$. This is a fixed-coefficient model and can easily be estimated by the ordinary least-squares (OLS) method. Denote the GM stock return and the market return by gm and sp, respectively. The result follows:

```
> da=read.table(``m-gmsp-excess-9003.txt'',header=F)
> gm=da[,1]
> sp=da[,2]
> fit=OLS(gm~sp)
> summary(fit)
Call:
OLS(formula = gm~sp)
Coefficients:
             Value Std. Error t value Pr(>|t|)
(Intercept) 0.1982 0.6302     0.3145  0.7535
         sp 1.0457 0.1453     7.1962  0.0000

Regression Diagnostics:
```

```
         R-Squared 0.2378
Adjusted R-Squared 0.2332
Durbin-Watson Stat 2.0290

Residual Diagnostics:
               Stat P-Value
Jarque-Bera  2.5348  0.2816
  Ljung-Box 24.2132  0.3362

Residual standard error: 8.13 on 166 degrees of freedom
```

Thus, the fitted model is

$$r_t = 0.20 + 1.0457 r_{M,t} + e_t, \qquad \hat{\sigma}_e = 8.13.$$

Based on the residual diagnostics, the model appears to be adequate for the GM stock returns with adjusted $R^2 = 23.3\%$.

As shown in Section 11.3, model (11.86) is a special case of the state-space model. We then estimate the model using SsfPack. The result is as follows:

```
> reg.m=function(parm,mX=NULL){
+ parm=exp(parm)   % log(sigma.e) is used to ensure
positiveness.
+ ssf.reg=GetSsfReg(mX)
+ ssf.reg$mOmega[3,3]=parm[1]
+ CheckSsf(ssf.reg)
+ }
> c.start=c(10)
> X.mtx=cbind(rep(1,168),sp)
> reg.fit=SsfFit(c.start,gm,"reg.m",mX=X.mtx)
RELATIVE FUNCTION CONVERGENCE
> names(reg.fit)
 [1] "parameters" "objective" "message" "grad.norm"
"iterations"
 [6] "f.evals"    "g.evals"   "hessian" "scale"     "aux"
[11] "call"       "vcov"
> sqrt(exp(reg.fit$parameters))
[1] 8.130114
> ssf.reg$mOmega[3,3]=exp(reg.fit$parameters)
> reg.s=SsfMomentEst(gm,ssf.reg,task="STSMO")
> reg.s$state.moment[10,]
   state.1  state.2
 0.1982025 1.045702
> sqrt(reg.s$state.variance[10,])
   state.1   state.2
 0.6302091 0.1453139
```

As expected, the result is in total agreement with that of the OLS method.

Finally, we entertain the time-varying CAPM of Section 11.3.1. The estimation result, including time plot of the smoothed response variable, is given below. The command `SsfCondDens` is used to compute the smoothed estimates of the state vector and observation without variance estimation.

```
> tv.capm =function(parm,mX=NULL){ %setup model for estimation
+ parm=exp(parm)    %parameterize in log for positiveness.
+ Phi.t = rbind(diag(2),rep(0,2))
+ Omega=diag(parm)
+ JPhi=matrix(-1,3,2)
+ JPhi[3,1]=1
+ JPhi[3,2]=2
+ Sigma=-Phi.t
+ ssf.tv=list(mPhi=Phi.t,
+ mOmega=Omega,
+ mJPhi=JPhi,
+ mSigma=Sigma,
+ mX=mX)
+ CheckSsf(ssf.tv)
+ }
> tv.start=c(0,0,0)   % starting values
> tv.mle=SsfFit(tv.start,gm,"tv.capm",mX=X.mtx)  % estimation
> sigma.mle=sqrt(exp(tv.mle$parameters))
> sigma.mle
[1] 4.907845e-05 1.219885e-02 8.125213e+00
% Smoothing
> smoEst.tv=SsfCondDens(gm,tv.capm(tv.mle$parameters,mX=X.
mtx),
+ task="STSMO")
> names(smoEst.tv)
[1] "state"    "response" "task"
> par(mfcol=c(2,2))  % plotting
> plot(gm,type='l',ylab='excess return')
> title(main='(a) Monthly simple excess returns')
> plot(smoEst.tv$response,type='l',ylab='rtn')
> title(main='(b) Expected returns')
> plot(smoEst.tv$state[,1],type='l',ylab='value')
> title(main='(c) Alpha(t)')
> plot(smoEst.tv$state[,2],type='l',ylab='value')
> title(main='(d) Beta(t)')
```

Note that estimates of σ_η and σ_ε are 4.91×10^{-5} and 1.22×10^{-2}, respectively. These estimates are close to zero, indicating that α_t and β_t of the time-varying market model are essentially constant for the GM stock returns. This is in agreement with the fact that the fixed-coefficient market model fits the data well. Figure 11.5 shows some plots for the time-varying CAPM fit. Part (a) is the monthly simple excess returns of GM stock from January 1990 to December 2003. Part (b) is the expected returns of GM stock, that is, $r_{t|T}$, where $T = 168$ is the sample size. Parts

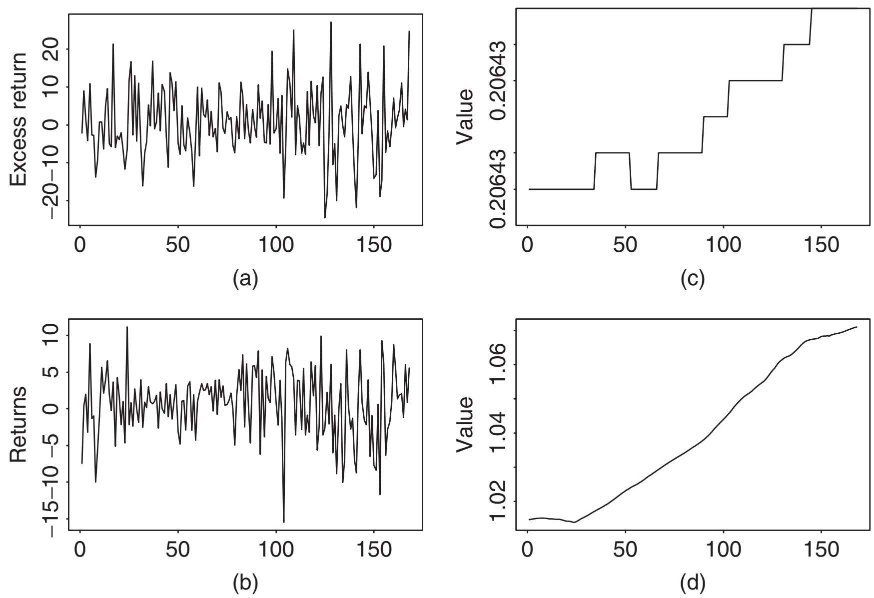


Figure 11.5 Time plots of some statistics for time-varying CAPM applied to monthly simple excess returns of General Motors stock. S&P 500 composite index return is used as market return: (a) monthly simple excess return, (b) expected returns $\boldsymbol{r}_{t|T}$, (c) α_t estimate, and (d) β_t estimate.

(c) and (d) are the time plots of the estimates of α_t and β_t. Given the tightness in the vertical scale, these two time plots confirm the assertion that a fixed-coefficient market model is adequate for the monthly GM stock return.

Example 11.3. In this example we reanalyze the series of quarterly earnings per share of Johnson & Johnson from 1960 to 1980 using the unobserved component model; see Chapter 2 for details of the data. The model considered is

$$y_t = \mu_t + \gamma_t + e_t, \qquad e_t \sim N(0, \sigma_e^2), \tag{11.87}$$

where y_t is the logarithm of the observed earnings per share, μ_t is the local trend component satisfying

$$\mu_{t+1} = \mu_t + \eta_t, \qquad \eta_t \sim N(0, \sigma_\eta^2),$$

and γ_t is the seasonal component that satisfies

$$(1 + B + B^2 + B^3)\gamma_t = \omega_t, \qquad \omega_t \sim N(0, \sigma_\omega^2),$$

that is, $\gamma_t = -\sum_{j=1}^{3} \gamma_{t-j} + \omega_t$. This model has three parameters—σ_e, σ_η, and σ_ω—and is a simple unobserved component model. It can be put in a state-space form as

$$\begin{bmatrix} \mu_{t+1} \\ \gamma_{t+1} \\ \gamma_t \\ \gamma_{t-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & -1 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_t \\ \gamma_t \\ \gamma_{t-1} \\ \gamma_{t-2} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \eta_t \\ \omega_t \end{bmatrix},$$

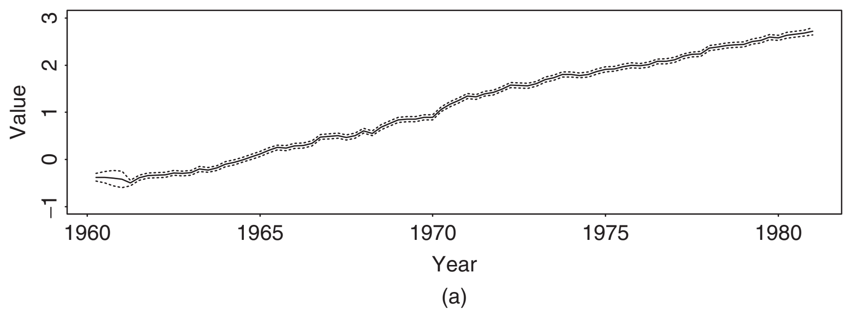
where the covariance matrix of $(\eta_t, \omega_t)'$ is $\text{diag}\{\sigma_\eta^2, \sigma_\omega^2\}$, and $y_t = [1, 1, 0, 0]\boldsymbol{s}_t + e_t$; see Section 11.3. This is a special case of the structural time series in SsfPack and can easily be specified using the command `GetSsfStsm`. Performing the maximum-likelihood estimation, we obtain $(\hat{\sigma}_e, \hat{\sigma}_\eta, \hat{\sigma}_\omega) = (2.04 \times 10^{-6}, 7.27 \times 10^{-2}, 2.93 \times 10^{-2})$.

```
> jnj=scan(file='q-jnj.txt')
> y=log(jnj)
% Estimation
> jnj.m=function(parm){
+ parm=exp(parm)
+ jnj.sea=GetSsfStsm(irregular=parm[1],level=parm[2],
+ seasonalDummy=c(parm[3],4))
+ CheckSsf(jnj.sea)
+ }
>
> c.start=c(0,0,0)  % Starting values
> jnj.est=SsfFit(c.start,y,"jnj.m")
> names(jnj.est)
 [1] "parameters" "objective" "message"  "grad.norm" "itera-
tions"
 [6] "f.evals"    "g.evals"   "hessian"  "scale"     "aux"
[11] "call"
> jnjest=exp(jnj.est$parameters)
> jnjest     % estimates
[1] 2.044516e-06 7.269655e-02 2.931691e-02
> jnj.ssf=GetSsfStsm(irregular=jnjest[1],level=jnjest[2],
+ seasonalDummy=c(jnjest[3],4)) % specify the model with esti-
mates
> CheckSsf(jnj.ssf)
$mPhi:
     [,1] [,2] [,3] [,4]
[1,]    1    0    0    0
[2,]    0   -1   -1   -1
[3,]    0    1    0    0
[4,]    0    0    1    0
[5,]    1    1    0    0
```

```
$mOmega:
            [,1]        [,2] [,3] [,4]         [,5]
[1,] 0.005284788 0.000000000    0    0 0.000000e+00
[2,] 0.000000000 0.000859481    0    0 0.000000e+00
[3,] 0.000000000 0.000000000    0    0 0.000000e+00
[4,] 0.000000000 0.000000000    0    0 0.000000e+00
[5,] 0.000000000 0.000000000    0    0 4.180047e-12
$mSigma:
     [,1] [,2] [,3] [,4]
[1,]   -1    0    0    0
[2,]    0   -1    0    0
[3,]    0    0   -1    0
[4,]    0    0    0   -1
[5,]    0    0    0    0
$mDelta:
     [,1]
[1,]    0
[2,]    0
[3,]    0
[4,]    0
[5,]    0
$mJPhi:
[1] 0
$mJOmega:
[1] 0
$mJDelta:
[1] 0
$mX:
[1] 0
$cT:
[1] 0
$cX:
[1] 0
$cY:
[1] 1
$cSt:
[1] 4
attr(, "class"):
[1] "ssf"   % below: smoothed components
> jnj.smo=SsfMomentEst(y,jnj.ssf,task="STSMO")
> up1=jnj.smo$state.moment[,1]+
+ 2*sqrt(jnj.smo$state.variance[,1])
> lw1=jnj.smo$state.moment[,1]-
+ 2*sqrt(jnj.smo$state.variance[,1])
> max(up1)  % obtain the range for plotting
```

```
[1] 2.795702
> min(lw1)
[1] -0.5948943
> up=jnj.smo$state.moment[,2]+
+ 2*sqrt(jnj.smo$state.variance[,2])
> lw=jnj.smo$state.moment[,2]-
+ 2*sqrt(jnj.smo$state.variance[,2])
> max(up)
[1] 0.3788652
> min(lw)
[1] -0.3552441
> par(mfcol=c(2,1)) % plotting
> plot(tdx,jnj.smo$state.moment[,1],type='l',xlab='year',
+ ylab='value',ylim=c(-1,3))
> lines(tdx,up1,lty=2)
> lines(tdx,lw1,lty=2)
> title(main='(a) Trend component')
> plot(tdx,jnj.smo$state.moment[,2],type='l',xlab='year',
+ ylab='value',ylim=c(-.5,.5))
> lines(tdx,up,lty=2)
> lines(tdx,lw,lty=2)
> title(main='(b) Seasonal component')
% Filtering and smoothing
> jnj.fil=KalmanFil(y,jnj.ssf,task="STFIL")
> jnj.smo1=KalmanSmo(jnj.fil,jnj.ssf)
> plot(tdx,jnj.fil$mOut[,1],type='l',xlab='year',ylab='resi')
> title(main='(a) 1-Step forecast error')
> plot(tdx,jnj.smo1$response.residuals[2:85],type='l',
+ xlab='year',ylab='resi')
> title(main='(b) Smoothing residual')
```

Figure 11.6 shows the smoothed estimates of the trend and seasonal components, that is, $\mu_{t|T}$ and $\gamma_{t|T}$ with $T = 84$, of the data. Of particular interest is that the seasonal pattern seems to evolve over time. Also shown are 95% pointwise confidence regions of the unobserved components. Figure 11.7 shows the residual plots, where part (a) gives the 1-step-ahead forecast errors computed by Kalman filter and part (b) is the smoothed response residuals of the fitted model. Thus, state-space modeling provides an alternative approach for analyzing seasonal time series. It should be noted that the estimated components in Figure 11.6 are not unique. They depend on the model specified and constraints used. In fact, there are infinitely many ways to decompose an observed time series into unobserved components. For instance, one can use a different specification for the seasonal component, for example, `sesonalTrig` in SsfPack, to obtain another decomposition for the earnings series of Johnson & Johnson. Thus, care must be exercised in interpreting the estimated components. However, for forecasting purposes, the choice of decomposition does not matter provided that the chosen one is a valid decomposition.



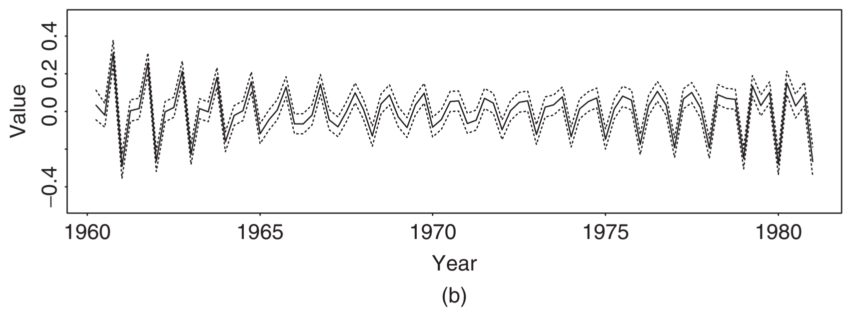


Figure 11.6 Smoothed components of fitting model (11.87) to logarithm of quarterly earnings per share of Johnson & Johnson from 1960 to 1980: (a) trend component and (b) seasonal component. Dotted lines indicate pointwise 95% confidence regions.

EXERCISES

11.1. Consider the ARMA(1,1) model $y_t - 0.8y_{t-1} = a_t + 0.4a_{t-1}$ with $a_t \sim N(0, 0.49)$. Convert the model into a state-space form using (a) Akaike's method, (b) Harvey's approach, and (c) Aoki's approach.

11.2. The file `aa-rv-20m.txt` contains the realized daily volatility series of Alcoa stock returns from January 2, 2003, to May 7, 2004; see the example in Section 11.1. The volatility series is constructed using 20-minute intradaily log returns.

(a) Fit an ARIMA(0,1,1) model to the log volatility series and write down the model.

(b) Estimate the local trend model in Eqs. (11.1) and (11.2) for the log volatility series. What are the estimates of σ_e and σ_η? Obtain time plots for the filtered and smoothed state variables with pointwise 95% confidence interval.

11.3. Consider the monthly simple excess returns of Pfizer stock and the S&P 500 composite index from January 1990 to December 2003. The excess returns are in `m-pfesp-ex9003.txt` with Pfizer stock returns in the first column.

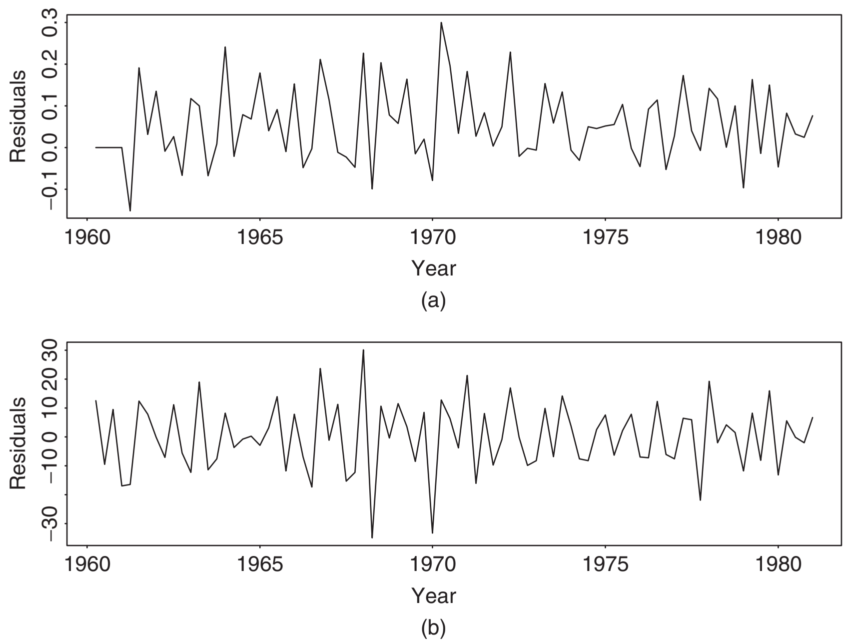


Figure 11.7 Residual series of fitting model (11.87) to logarithm of quarterly earnings per share of Johnson & Johnson from 1960 to 1980: (a) 1-step-ahead forecast error v_t and (b) smoothed residuals of response variable.

(a) Fit a fixed-coefficient market model to the Pfizer stock return. Write down the fitted model.

(b) Fit a time-varying CAPM to the Pfizer stock return. What are the estimated standard errors of the innovations to the α_t and β_t series? Obtain time plots of the smoothed estimates of α_t and β_t.

11.4. Consider the AR(3) model

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \phi_3 x_{t-3} + a_t, \qquad a_t \sim N(0, \sigma_a^2),$$

and suppose that the observed data are

$$y_t = x_t + e_t, \qquad e_t \sim N(0, \sigma_e^2),$$

where $\{e_t\}$ and $\{a_t\}$ are independent and the initial values of x_j with $j \leq 0$ are independent of e_t and a_t for $t > 0$.

(a) Convert the model into a state-space form.

(b) If $E(e_t) = c$, which is not zero, what is the corresponding state-space form for the system?

11.5. The file `m-ppiaco4709.txt` contains year, month, day, and U.S. producer price index (PPI) from January 1947 to November 2009. The index is for all commodities and not seasonally adjusted. Let $z_t = \ln(Z_t) - \ln(Z_{t-1})$, where Z_t is the observed monthly PPI. It turns out that an AR(3) model is adequate for z_t if the minor seasonal dependence is ignored. Let y_t be the sample mean-corrected series of z_t.

(a) Fit an AR(3) model to y_t and write down the fitted model.

(b) Suppose that y_t has independent measurement errors so that $y_t = x_t + e_t$, where x_t is a zero-mean AR(3) process and $\text{Var}(e_t) = \sigma_e^2$. Use a state-space form to estimate parameters, including the innovational variances to the state and σ_e^2. Write down the fitted model and obtain a time plot of the smoothed estimate of x_t. Also, show the time plot of filtered response residuals of the fitted state-space model.

REFERENCES

Akaike, H. (1975). Markovian representation of stochastic processes by canonical variables. *SIAM Journal on Control* **13**: 162–173.

Aoki, M. (1987). *State Space Modeling of Time Series*. Springer, New York.

Anderson, B. D. O. and Moore, J. B. (1979). *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, NJ.

Chan, N. H. (2002). *Time Series: Applications to Finance*. Wiley, Hoboken, NJ.

de Jong, P. (1989). Smoothing and interpolation with the state space model. *Journal of the American Statistical Association* **84**: 1085–1088.

Durbin, J. and Koopman, S. J. (2001). *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford, UK.

Hamilton, J. (1994). *Time Series Analysis*. Princeton University Press, Princeton, NJ.

Harvey, A. C. (1993). *Time Series Models*, 2nd ed. Harvester Wheatsheaf, Hemel Hempstead, UK.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, Transactions ASMA, Series D* **82**: 35–45.

Kim, C. J. and Nelson, C. R. (1999). *State Space Models with Regime Switching*. Academic, New York.

Kitagawa, G. and Gersch, W. (1996). *Smoothness Priors Analysis of Time Series*. Springer, New York.

Koopman, S. J. (1993). Disturbance smoother for state space models. *Biometrika* **80**: 117–126.

Koopman, S. J., Shephard, N., and Doornik, J. A. (1999). Statistical algorithms for models in state-space form using *SsfPack* 2.2. *Econometrics Journal* **2**: 113–166. Also available at http://www.ssfpack.com/.

Shumway, R. H. and Stoffer, D. S. (2000). *Time Series Analysis and Its Applications*. Springer, New York.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*, 2nd ed. Springer, New York.