

## Lecture 9: State-Space Models and Kalman Filters

The state-space model provides a flexible approach to time series analysis, especially for ease in estimation and in 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 state-space approach to time series analysis before introducing the general state-space model.

This handout is based on Chapter 11 of Tsay (2005, Wiley). It is intended for use in my lecture only. Don't distributed without a written permission from the publisher John Wiley and Sons, Inc.

### 9.1 Local Trend Model

Consider the univariate time series  $y_t$  satisfying

$$y_t = \mu_t + e_t, \quad e_t \sim N(0, \sigma_e^2) \quad (9.1)$$

$$\mu_{t+1} = \mu_t + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2) \quad (9.2)$$

where  $\{e_t\}$  and  $\{\eta_t\}$  are two independent Gaussian white noise series and  $t = 1, \dots, T$ . The initial value  $\mu_1$  is either given or follows a known distribution, and it is independent of  $\{e_t\}$  and  $\{\eta_t\}$  for  $t > 0$ . Here  $\mu_t$  is a pure *random walk* with initial value  $\mu_1$  and  $y_t$  is an observed version of  $\mu_t$  with added noise  $a_t$ . In the literature,  $\mu_t$  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  $\mu_t$  because  $\{e_t\}$  is not serially correlated.

**Example 1.** To illustrate the ideas of the state-space model and Kalman filter, we consider the intra-daily 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. The series used in the demonstration is the logarithm of the daily realized volatility.

Figure 9.1 shows the time plot of the logarithms of the realized volatility of Alcoa stock from January 02, 2003 to May 7, 2004. The transactions data are obtained from the TAQ database of NYSE. If ARIMA models are entertained, we obtain an ARIMA(0,1,1) model

$$(1 - B)y_t = (1 - 0.855B)a_t, \quad \hat{\sigma}_a = 0.5184 \quad (9.3)$$

where  $y_t$  is the log realized volatility, and the standard error of  $\hat{\theta}$  is 0.029. The residuals show  $Q(12) = 12.4$  with p-value 0.33, indicating that there is no significant serial correlation in the residuals. Similarly, the squared residuals give  $Q(12) = 8.2$  with p-value 0.77, suggesting no ARCH effects in the series.

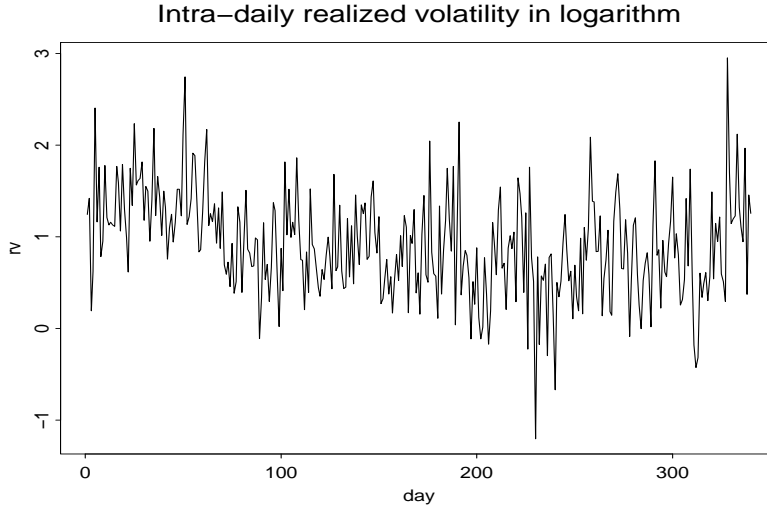


Figure 9.1: Time plot of the logarithms of intra-daily realized volatility of the Alcoa stock from January 2, 2003 to May 7, 2004. The realized volatility is computed from the intraday 10-minute log returns measured in percentage.

Since  $\hat{\theta}$  is positive, we can transform the ARIMA(0,1,1) model into a local trend model in Eqs. (9.1)-(9.2). The maximum likelihood estimates 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 Subsection 9.1.6. Here we treat the two estimates as given and use the model to demonstrate the application of Kalman filter.

### 9.1.1 Statistical Inference

Return to the state-space model in Eqs. (9.1)-(9.2). The object of the analysis is to infer properties of the state  $\mu_t$  from the data  $\{y_t | t = 1, \dots, T\}$  and the model. Three types of inference are commonly discussed in the literature. They are *filtering*, *prediction*, and *smoothing*. Let  $F_t = \{y_1, \dots, y_t\}$  be the information available at time  $t$  (inclusive) and assume that the model is known, including all parameters. The three types of inference can be briefly described as follows:

- *filtering*: Filtering means to recover the state variable  $\mu_t$  given  $F_t$ , i.e. to remove the measurement errors from the data.
- *Prediction*: Prediction means to forecast  $\mu_{t+h}$  or  $y_{t+h}$  for  $h > 0$  given  $F_t$ , where  $t$  is the forecast origin.
- *Smoothing*: Smoothing is to estimate  $\mu_t$  given  $F_T$ , where  $T > t$ .

To describe the inference more precisely, we introduce some notation. Let  $\mu_{t|j} = E(\mu_t | F_j)$  be the conditional expectation of  $\mu_t$  given the data  $F_j$ , and  $\Sigma_{t|j} = \text{Var}(\mu_t | F_j)$  be the conditional variance of  $\mu_t$  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 one-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. (9.1),

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

Consequently,

$$v_t = y_t - y_{t|t-1} = y_t - \mu_{t|t-1} \quad (9.4)$$

and

$$\begin{aligned} V_t &= \text{Var}(y_t - \mu_{t|t-1}|F_{t-1}) = \text{Var}(\mu_t + e_t - \mu_{t|t-1}|F_{t-1}) \\ &= \text{Var}(\mu_t - \mu_{t|t-1}|F_{t-1}) + \text{Var}(e_t|F_{t-1}) = \Sigma_{t|t-1} + \sigma_e^2. \end{aligned} \quad (9.5)$$

It is also easy to see that

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

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

Thus, as expected, the one-step ahead forecast error is uncorrelated (hence, independent) with  $y_j$  for  $j < t$ . Furthermore, for the linear model in Eqs. (9.1)-(9.2),  $\mu_{t|t} = E(\mu_t|F_t) = E(\mu_t|F_{t-1}, v_t)$  and  $\Sigma_{t|t} = \text{Var}(\mu_t|F_t) = \text{Var}(\mu_t|F_{t-1}, v_t)$ . 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 Kalman filter under normality. They can be shown via multivariate linear regression method or factorization of the joint density.

**Theorem 1** Suppose that  $\mathbf{x}$ ,  $\mathbf{y}$ , and  $\mathbf{z}$  are three random vectors such that their joint distribution is multivariate normal with mean  $E(\mathbf{w}) = \boldsymbol{\mu}_w$  and covariance matrix  $\boldsymbol{\Sigma}_{mw} = \text{Cov}(\mathbf{m}, \mathbf{w})$ , where  $w$  and  $m$  are  $x, y$ , and  $z$ . In addition, assume that the diagonal block covariance matrix  $\boldsymbol{\Sigma}_{ww}$  is non-singular for  $w = x, y, z$ , and  $\boldsymbol{\Sigma}_{yz} = \mathbf{0}$ . Then,

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

### 9.1.2 Kalman Filter

The goal of *Kalman filter* is to update the knowledge of the state variable recursively when a new data point becomes available. That is, we like to obtain the conditional distribution of  $\mu_t$  given  $F_t$  based on the new data  $y_t$  and the conditional distribution of  $\mu_t$  given  $F_{t-1}$ . 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 Kalman filter, it suffices to consider the joint conditional distribution of  $(\mu_t, v_t)'$  given  $F_{t-1}$  before applying Theorem 1.

The conditional distribution of  $v_t$  given  $F_{t-1}$  is normal with mean zero and variance given in Eq. (9.5), and that of  $\mu_t$  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  $\mu_t$  and  $v_t$  given  $F_{t-1}$ . From the definition,

$$\begin{aligned} \text{Cov}(\mu_t, v_t|F_{t-1}) &= E(\mu_t v_t|F_{t-1}) = E[\mu_t(y_t - \mu_{t|t-1})|F_{t-1}] \quad [\text{by Eq.(9.4)}] \\ &= E[\mu_t(\mu_t + e_t - \mu_{t|t-1})|F_{t-1}] \\ &= E[\mu_t(\mu_t - \mu_{t|t-1})|F_{t-1}] + E(\mu_t e_t|F_{t-1}) \\ &= E[(\mu_t - \mu_{t|t-1})^2|F_{t-1}] = \text{Var}(\mu_t|F_{t-1}) = \Sigma_{t|t-1}, \end{aligned} \quad (9.6)$$

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 that

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

By Theorem 1, the conditional distribution of  $\mu_t$  given  $F_t$  is normal with mean and variance as

$$\mu_{t|t} = \mu_{t|t-1} + \frac{\Sigma_{t|t-1} v_t}{V_t} = \mu_{t|t-1} + K_t v_t \quad (9.7)$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} + \frac{\Sigma_{t|t-1}^2}{V_t} = \Sigma_{t|t-1}(1 - K_t), \quad (9.8)$$

where  $K_t = \Sigma_{t|t-1}/V_t$  is commonly referred to as the *Kalman gain*, which is the regression coefficient of  $\mu_t$  on  $v_t$ . In Kalman filter, Kalman gain is the factor that governs the contribution of the new shock  $v_t$  to the state variable  $\mu_{t+1}$ .

Next, one can make use of the knowledge of  $\mu_t$  given  $F_t$  to predict  $\mu_{t+1}$  via Eq. (9.2). Specifically, we have

$$\mu_{t+1|t} = E(\mu_{t+1}|F_t) = E(\mu_t|F_t) = \mu_{t|t}, \quad (9.9)$$

$$\Sigma_{t+1|t} = \text{Var}(\mu_{t+1}|F_t) = \text{Var}(\mu_t|F_t) + \text{Var}(\eta_t) = \Sigma_{t|t} + \sigma_\eta^2. \quad (9.10)$$

Once the new data  $y_{t+1}$  is observed, one can repeat the above procedure to update the knowledge of  $\mu_{t+1}$ . This is the famous *Kalman Filter* algorithm; see Kalman (1960).

In summary, putting Eqs. (9.4)-(9.10) together and given the initial condition that  $\mu_1$  is distributed as  $N(\mu_{1|0}, \Sigma_{1|0})$ , the Kalman filter for the local trend model is as follows:

$$\begin{aligned} v_t &= y_t - \mu_{t|t-1} \\ V_t &= \Sigma_{t|t-1} + \sigma_e^2 \\ K_t &= \Sigma_{t|t-1}/V_t \\ \mu_{t+1|t} &= \mu_{t|t-1} + K_t v_t \\ \Sigma_{t+1|t} &= \Sigma_{t|t-1}(1 - K_t) + \sigma_\eta^2, \quad t = 1, \dots, T. \end{aligned} \quad (9.11)$$

There are many ways to derive the Kalman filter. We use Theorem 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. For the local-trend model in Eqs. (9.1)-(9.2), the two

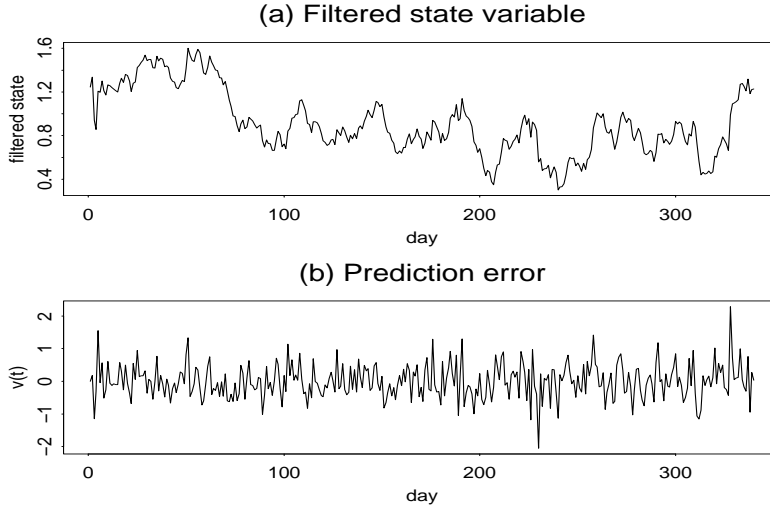


Figure 9.2: Time plots of the output of Kalman filter applied to the daily realized log volatility of Alcoa stock based on the local-trend state-space model: (a) The filtered state  $\mu_{t|t}$  and (b) the one-step ahead forecast error  $v_t$ .

parameters  $\sigma_e$  and  $\sigma_\eta$  can be estimated via maximum likelihood method. Again, Kalman filter is useful in evaluating the likelihood function of the data in estimation. We shall discuss estimation in Subsection 9.1.6.

**Example 1** (Continued). To illustrate the application of 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  $\mu_{1|0} = 0$ . Figure 9.2(a) shows the time plot of the filtered state variable  $\mu_{t|t}$  and Figure 9.2(b) is the time plot of the one-step ahead forecast error  $v_t$ . Compared with Figure 9.1, the filtered states are smoother. The forecast errors appear to be stable and center around zero. These forecast errors are out-of-sample one-step ahead prediction errors.

### 9.1.3 Properties of Forecast error

The one-step ahead forecast errors  $\{v_t\}$  are useful in many ways, and it pays to study 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  $\{y_1, \dots, y_t\}$  and  $\mu_{1|0}$ . Specifically, by repeated substitutions,

$$\begin{aligned} v_1 &= y_1 - \mu_{1|0} \\ v_2 &= y_2 - \mu_{2|1} = y_2 - \mu_{1|0} - K_1(y_1 - \mu_{1|0}) \\ v_3 &= y_3 - \mu_{3|2} = y_3 - \mu_{1|0} - K_2(y_2 - \mu_{1|0}) - K_1(1 - K_2)(y_1 - \mu_{1|0}), \end{aligned}$$

and so on. In matrix form, the prior transformation is

$$\mathbf{v} = \mathbf{K}(\mathbf{y} - \mathbf{1}_T \mu_{1|0}), \quad (9.12)$$

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

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

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  $i = 2, \dots, T$  and  $j = 1, \dots, i - 2$ . Also, from the definition,  $K_t$  does not depend on  $\mu_{1|0}$  or the data  $\{y_1, \dots, y_t\}$ ; it depends on  $\Sigma_{1|0}$  and  $\sigma_e^2$  and  $\sigma_\eta^2$ .

The transformation in Eq. (9.12) 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

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

Eq. (9.12) indicates that the transformation from  $y_t$  to  $v_t$  has a unity Jacobian so that  $p(\mathbf{v}) = p(\mathbf{y})$ . Furthermore, since  $\mu_{1|0}$  is given,  $p(v_1) = p(y_1)$ . Consequently, the joint probability density function of  $\mathbf{v}$  is

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

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

Second, the Kalman filter provides a Cholesky decomposition of the covariance matrix of  $\mathbf{y}$ . To see this, let  $\mathbf{\Omega} = \text{Cov}(\mathbf{y})$ . Eq. (9.12) shows that  $\text{Cov}(\mathbf{v}) = \mathbf{K}\mathbf{\Omega}\mathbf{K}'$ . On the other hand,  $\{v_t\}$  are mutually independent with  $\text{Var}(v_t) = V_t$ . Therefore,  $\mathbf{K}\mathbf{\Omega}\mathbf{K}' = \text{diag}\{V_1, \dots, V_T\}$ , which is precisely a Cholesky decomposition of  $\mathbf{\Omega}$ . The elements  $k_{ij}$  of the matrix  $\mathbf{K}$  thus have some nice interpretations.

### *State error recursion*

Turn to the estimation error of state variable  $\mu_t$ . Define

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

as the forecast error of the state variable  $\mu_t$  given data  $F_{t-1}$ . From Subsection 9.1.1,  $\text{Var}(x_t | F_{t-1}) = \Sigma_{t|t-1}$ . From the Kalman filter in Eq. (9.11),

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

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

$$v_t = x_t + e_t, \quad x_{t+1} = L_t x_t + \eta_t - K_t e_t, \quad t = 1, \dots, T, \quad (9.13)$$

where  $x_1 = \mu_1 - \mu_{1|0}$ . Eq. (9.13) is in the form of a time-varying state-space model with  $x_t$  as the state variable and  $v_t$  as the observation.

### 9.1.4 State Smoothing

Next we consider the estimation of the state variables  $\{\mu_1, \dots, \mu_T\}$  given the data  $F_T$  and the model. That is, given the state-space model in Eqs. (9.1)-(9.2), 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:

- (a) All distributions involved are normal so that we can write the conditional distribution of  $\mu_t$  given  $F_T$  as  $N(\mu_{t|T}, \Sigma_{t|T})$ , where  $t \leq T$ . We refer to  $\mu_{t|T}$  as the *smoothed state* at time  $t$  and  $\Sigma_{t|T}$  as the *smoothed state variance*.
- (b) Based on the properties of  $\{v_t\}$  discussed in Subsection 9.1.3,  $\{v_1, \dots, v_T\}$  are mutually independent and are linear functions of  $\{y_1, \dots, y_T\}$ .
- (c) If  $y_1, \dots, y_T$  are fixed, then  $F_{t-1}$  and  $\{v_t, \dots, v_T\}$  are fixed, and vice versa.
- (d)  $\{v_t, \dots, v_T\}$  are independent of  $F_{t-1}$  with mean zero and variance  $\text{Var}(v_j) = V_j$  for  $j \geq t$ .

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

$$\begin{aligned}
\mu_{t|T} &= E(\mu_t|F_T) = E(\mu_t|F_{t-1}, v_t, \dots, v_T) \\
&= E(\mu_t|F_{t-1}) + \text{Cov}[\mu_t, (v_t, \dots, v_T)'] \text{Cov}[(v_t, \dots, v_T)']^{-1} (v_t, \dots, v_T)' \\
&= \mu_{t|t-1} + \begin{bmatrix} \text{Cov}(\mu_t, v_t) \\ \text{Cov}(\mu_t, v_{t+1}) \\ \vdots \\ \text{Cov}(\mu_t, v_T) \end{bmatrix}' \begin{bmatrix} V_t & 0 & \cdots & 0 \\ 0 & V_{t+1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & V_T \end{bmatrix}^{-1} \begin{bmatrix} v_t \\ v_{t+1} \\ \vdots \\ v_T \end{bmatrix} \\
&= \mu_{t|t-1} + \sum_{j=t}^T \text{Cov}(\mu_t, v_j) V_j^{-1} v_j. \tag{9.14}
\end{aligned}$$

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

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

Similarly, we have

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

Consequently, Eq.(9.14) becomes

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

where

$$q_{t-1} = \frac{v_t}{V_t} + L_t \frac{v_{t+1}}{V_{t+1}} + L_t L_{t+1} \frac{v_{t+2}}{V_{t+2}} + \cdots + \left( \prod_{j=t}^{T-1} L_j \right) \frac{v_T}{V_T} \quad (9.15)$$

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

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

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

$$q_{t-1} = \frac{v_t}{V_t} + L_t q_t, \quad t = T, T-1, \dots, 1. \quad (9.16)$$

Putting Eqs.(9.14)-(9.16) together, we have a backward recursive algorithm to compute the smoothed state variables:

$$q_{t-1} = V_t^{-1} v_t + L_t q_t, \quad \mu_{t|T} = \mu_{t|t-1} + \Sigma_{t|t-1} q_{t-1}, \quad t = T, \dots, 1, \quad (9.17)$$

where  $q_T = 0$ , and  $\mu_{t|t-1}$ ,  $\Sigma_{t|t-1}$  and  $L_t$  are available from the Kalman filter in Eq. (9.11).

### ***Smoothed state variance***

The variance of the smoothed state variable  $\mu_{t|T}$  can be derived in a similar manner via Theorem 1(4). Specifically, letting  $\mathbf{v}_t^T = (v_t, \dots, v_T)'$ , we have

$$\begin{aligned} \Sigma_{t|T} &= \text{Var}(\mu_t | F_T) = \text{Var}(\mu_t | F_{t-1}, v_t, \dots, v_T) \\ &= \text{Var}(\mu_t | F_{t-1}) - \text{Cov}[\mu_t, (\mathbf{v}_t^T)'] \text{Cov}[(\mathbf{v}_t^T)']^{-1} \text{Cov}[\mu_t, (\mathbf{v}_t^T)'] \\ &= \Sigma_{t|t-1} + \sum_{j=t}^T [\text{Cov}(\mu_t, v_j)]^2 V_j^{-1}, \end{aligned} \quad (9.18)$$

where  $\text{Cov}(\mu_t, v_j) = \text{Cov}(x_t, v_j)$  are given earlier after Eq. (9.14). Thus,

$$\begin{aligned} \Sigma_{t|T} &= \Sigma_{t|t-1} - \Sigma_{t|t-1}^2 \frac{1}{V_t} - \Sigma_{t|t-1}^2 L_t^2 \frac{1}{V_{t+1}} - \cdots - \Sigma_{t|t-1}^2 \left( \prod_{j=t}^{T-1} L_j^2 \right) \frac{1}{V_T} \\ &= \Sigma_{t|t-1} - \Sigma_{t|t-1}^2 M_{t-1}, \end{aligned} \quad (9.19)$$

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 one-step ahead forecast errors after time  $t-1$ . Let  $M_T = 0$  because no one-step ahead forecast error is available after time index

$T$ . The statistic  $M_{t-1}$  can be written as

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

Note that from the independence of  $\{v_t\}$  and Eq. (9.15), we have

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

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

$$M_{t-1} = V_t^{-1} + L_t^2 M_t, \quad \Sigma_{t|T} = \Sigma_{t|t-1} - \Sigma_{t|t-1}^2 M_{t-1}, \quad t = T, \dots, 1, \quad (9.20)$$

where  $M_T = 0$ .

**Example 1** (Continued). Applying the Kalman filter and state-smoothing algorithms in Eqs. (9.17) and (9.20) 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 associated variances. Figure 9.3 shows the filtered state variable and its 95% pointwise confidence interval whereas Figure 9.4 provides the time plot of smoothed state variable and its 95% pointwise confidence interval. As expected, the smoothed state variable is smoother than the filtered state variable. The confidence intervals for the smoothed state variable 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  $\Sigma_{1|0}$ .

### 9.1.5 Missing Values

An advantage of state-space model is handling missing values. Suppose that the observations at  $t = \ell + 1, \dots, \ell + h$  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. From Eq. (9.2), we have that for  $t \in \{\ell + 1, \dots, \ell + h\}$ ,

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

where it is understood that the summation term is zero if its lower limit is greater than its upper limit. Therefore,

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

for  $\ell + 1 \leq t \leq \ell + h$ . Consequently, we have

$$\mu_{t|t-1} = \mu_{t-1|t-2}, \quad \Sigma_{t|t-1} = \Sigma_{t-1|t-2} + \sigma_\eta^2, \quad (9.21)$$

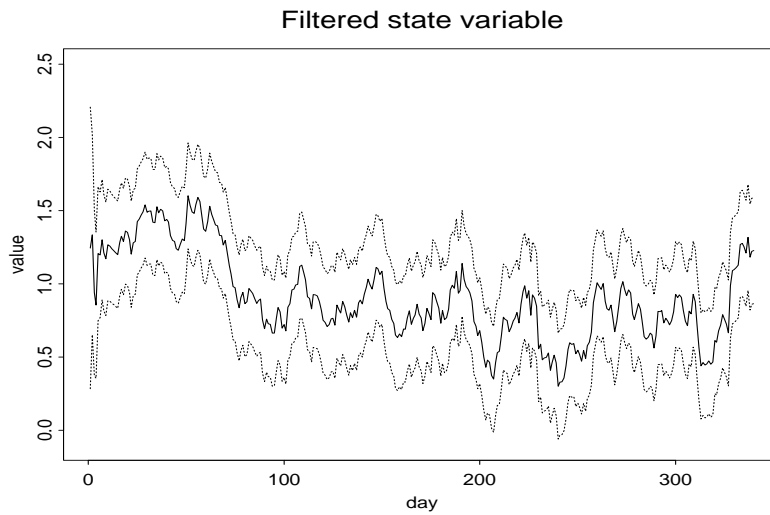


Figure 9.3: Filtered state variable  $\mu_{t|t}$  and its 95% pointwise confidence interval for the daily log realized volatility of Alcoa stock returns based on the fitted local-trend state-space model.

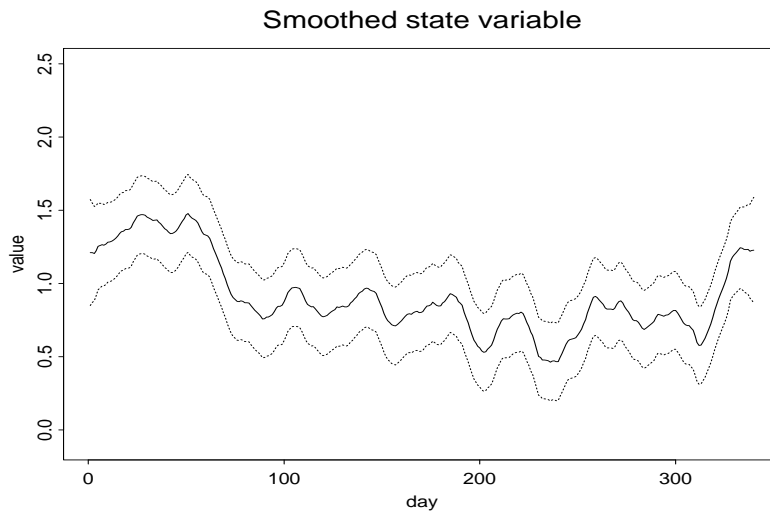


Figure 9.4: Smoothed state variable  $\mu_{t|T}$  and its 95% pointwise confidence interval for the daily log realized volatility of Alcoa stock returns based on the fitted local-trend state-space model.

for  $t = \ell + 1, \dots, \ell + h$ . These results show that we can continue to apply the Kalman filter algorithm in Eq. (9.11) by taking  $v_t = 0$  and  $K_t = 0$  for  $t = \ell + 1, \dots, \ell + h$ . 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  $K_t = 0$ .

### 9.1.6 Estimation

In this subsection, we consider the estimation of  $\sigma_e$  and  $\sigma_\eta$  of the local trend model in Eqs. (9.1)-(9.2). Based on properties of forecast errors discussed in Subsection 9.1.3, the Kalman filter provides an efficient way to evaluate the likelihood function of the data for estimation. Specifically, the likelihood function under normality is

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

where  $y_1 \sim N(\mu_{1|0}, V_1)$  and  $v_t = y_t - \mu_{t|t-1}$  is normally distributed with mean zero and variance  $V_t$ . Consequently, assuming  $\mu_{1|0}$  and  $\Sigma_{1|0}$  are known, and taking the logs, we have

$$\ln[L(\sigma_e, \sigma_\eta)] = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^T \left[ \ln(V_t) + \frac{v_t^2}{V_t} \right]. \quad (9.22)$$

Therefore, the log-likelihood function, including cases with missing values, can be evaluated recursively via Kalman filter. Many software packages perform state-space model estimation via Kalman filter algorithm. In this chapter, we use the *SsfPack* program developed by Koopman, Shephard, and Doornik (1999) and available in S-Plus.

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 Kalman filter to obtain the one-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.

## 9.2 Linear state-space Models

We now extend the local trend model to a 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

$$\mathbf{s}_{t+1} = \mathbf{d}_t + \mathbf{T}_t \mathbf{s}_t + \mathbf{R}_t \boldsymbol{\eta}_t \quad (9.23)$$

$$\mathbf{y}_t = \mathbf{c}_t + \mathbf{Z}_t \mathbf{s}_t + \mathbf{e}_t \quad (9.24)$$

where  $\mathbf{s}_t = (s_{1t}, \dots, s_{mt})'$  is an  $m$ -dimensional state vector,  $\mathbf{y}_t = (y_{1t}, \dots, y_{kt})'$  is a  $k$ -dimensional observation vector,  $\mathbf{d}_t$  and  $\mathbf{c}_t$  are  $m$ - and  $k$ -dimensional deterministic vectors,  $\mathbf{T}_t$  and  $\mathbf{Z}_t$  are  $m \times m$

and  $k \times m$  coefficient matrices,  $\mathbf{R}_t$  is an  $m \times n$  matrix often consisting of a subset of columns of the  $m \times m$  identity matrix,  $\{\boldsymbol{\eta}_t\}$  and  $\{\mathbf{e}_t\}$  are  $n$ - and  $k$ -dimensional Gaussian white noise series such that

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

where  $\mathbf{Q}_t$  and  $\mathbf{H}_t$  are positive-definite matrices. We assume that  $\{\mathbf{e}_t\}$  and  $\{\boldsymbol{\eta}_t\}$  are independent, but this condition can be relaxed if necessary. The initial state  $\mathbf{s}_1$  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  $\mathbf{e}_t$  and  $\boldsymbol{\eta}_t$  for  $t > 0$ .

Eq. (9.24) is the *measurement* or *observation* equation that relates the vector of observations  $\mathbf{y}_t$  to the state vector  $\mathbf{s}_t$ , the explanatory variable  $\mathbf{c}_t$  and the measurement error  $\mathbf{e}_t$ . Eq. (9.23) 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  $\mathbf{T}_t$ ,  $\mathbf{R}_t$ ,  $\mathbf{Q}_t$ ,  $\mathbf{Z}_t$ , and  $\mathbf{H}_t$  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 methods.

The state-space model in Eqs. (9.23)-(9.24) can be rewritten in a compact form as

$$\begin{bmatrix} \mathbf{s}_{t+1} \\ \mathbf{y}_t \end{bmatrix} = \boldsymbol{\delta}_t + \boldsymbol{\Phi}_t \mathbf{s}_t + \mathbf{u}_t, \quad (9.25)$$

where

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

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

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

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  $\lambda$  is a large real number, which can approach infinity.

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

### 9.3 Model Transformation

To appreciate the flexibility of state-space model, we rewrite ARMA models in the state-space form.

Consider a zero-mean ARMA( $p, q$ ) model

$$\phi(B)y_t = \theta(B)a_t, \quad a_t \sim N(0, \sigma_a^2), \quad (9.26)$$

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 an ARMA model into state-space form. We discuss three

methods. Let  $m = \max(p, q + 1)$  and rewrite the ARMA model in Eq. (9.26) as

$$y_t = \sum_{i=1}^m \phi_i y_{t-i} + a_t - \sum_{j=1}^{m-1} \theta_j a_{t-j}, \quad (9.27)$$

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  $\mathbf{s}_t$  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. (9.26) with  $m = \max(p, q + 1)$ ,  $\mathbf{s}_t = (y_{t|t}, y_{t+1|t}, \dots, y_{t+m-1|t})'$ , where  $y_{t+j|t} = E(y_{t+j}|F_t)$  is the conditional expectation of  $y_{t+j}$  given  $F_t = \{y_1, \dots, y_t\}$ . Since  $y_{t|t} = y_t$ , the first element of  $\mathbf{s}_t$  is  $y_t$ . Thus, the observation equation is

$$y_t = \mathbf{Z} \mathbf{s}_t, \quad (9.28)$$

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

$$s_{1,t+1} = y_{t+1} = y_{t+1|t} + (y_{t+1} - y_{t+1|t}) = s_{2t} + a_{t+1}. \quad (9.29)$$

Next, consider the MA representation of ARMA models

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

where  $\psi_0 = 1$  and other  $\psi$ -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

$$\begin{aligned} \psi_1 &= \phi_1 - \theta_1 \\ \psi_2 &= \phi_1 \psi_1 + \phi_2 - \theta_2 \\ &\vdots \\ \psi_{m-1} &= \phi_1 \psi_{m-2} + \phi_2 \psi_{m-3} + \dots + \phi_{m-2} \psi_1 + \phi_{m-1} - \theta_{m-1} \\ &= \sum_{i=1}^{m-1} \phi_i \psi_{m-1-i} - \theta_{m-1}. \end{aligned} \quad (9.30)$$

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

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

and

$$y_{t+j|t+1} = E(y_{t+j}|F_{t+1}) = \psi_{j-1} a_{t+1} + \psi_j a_t + \psi_{j+1} a_{t-1} + \dots = \psi_{j-1} a_{t+1} + y_{t+j|t}.$$

Thus, for  $j > 0$ , we have

$$y_{t+j|t+1} = y_{t+j|t} + \psi_{j-1} a_{t+1}. \quad (9.31)$$

This result is referred to as the forecast updating formula of ARMA models. It provides a simple way to update forecasts 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- $t$  forecasts are revised based on this new information with weights  $\psi_{j-1}$  to compute time- $(t+1)$  forecasts.

Finally, from Eq. (9.27) and using  $E(a_{t+j}|F_{t+1}) = 0$  for  $j > 1$ , we have

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

Using Eq. (9.31), the prior equation becomes

$$\begin{aligned} y_{t+m|t+1} &= \sum_{i=1}^{m-1} \phi_i (y_{t+m-i|t} + \psi_{m-i-1} a_{t+1}) + \psi_m y_{t|t} - \theta_{m-1} a_{t+1} \\ &= \sum_{i=1}^m \phi_i y_{t+m-i|t} + \left( \sum_{i=1}^{m-1} \phi_i \psi_{m-1-i} - \theta_{m-1} \right) a_{t+1} \\ &= \sum_{i=1}^m \phi_i y_{t+m-i|t} + \psi_{m-1} a_{t+1}, \end{aligned} \quad (9.32)$$

where the last equality uses Eq. (9.30). Putting Eqs. (9.29), (9.31) for  $j = 2, \dots, m-1$ , and (9.32) together, we have

$$\begin{bmatrix} y_{t+1} \\ y_{t+2|t+1} \\ \vdots \\ y_{t+m-1|t+1} \\ y_{t+m|t+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdot & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \cdot & 1 \\ \phi_m & \phi_{m-1} & \phi_{m-2} & \cdot & \phi_1 \end{bmatrix} \begin{bmatrix} y_t \\ y_{t+1|t} \\ \vdots \\ y_{t+m-2|t} \\ y_{t+m-1|t} \end{bmatrix} + \begin{bmatrix} 1 \\ \psi_1 \\ \vdots \\ \psi_{m-2} \\ \psi_{m-1} \end{bmatrix} a_{t+1}. \quad (9.33)$$

Thus, the transition equation of the Akaike's approach is

$$\mathbf{s}_{t+1} = \mathbf{T} \mathbf{s}_t + \mathbf{R} \eta_t, \quad \eta_t \sim N(0, \sigma_a^2), \quad (9.34)$$

where  $\eta_t = a_{t+1}$ , and  $\mathbf{T}$  and  $\mathbf{R}$  are the coefficient matrices in Eq. (9.33).

### **Harvey's approach**

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

$$\mathbf{s}_{t+1} = \mathbf{T} \mathbf{s}_t + \mathbf{R} \eta_t, \quad \eta_t \sim N(0, \sigma_a^2) \quad (9.35)$$

$$y_t = \mathbf{Z} \mathbf{s}_t \quad (9.36)$$

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

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

and  $\mathbf{d}_t$ ,  $\mathbf{c}_t$  and  $\mathbf{H}_t$  are all zero. The model in Eqs. (9.35)-(9.36) 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 ways to convert an ARMA model to state-space form. First, consider the MA model, i.e.  $y_t = \theta(B)a_t$ . In this case, we can simply define  $\mathbf{s}_t = (a_{t-q}, a_{t-q+2}, \dots, a_{t-1})'$  and obtain the state-space form

$$\begin{aligned} \begin{bmatrix} a_{t-q+1} \\ a_{t-q+2} \\ \vdots \\ a_{t-1} \\ a_t \end{bmatrix} &= \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} a_{t-q} \\ a_{t-q+1} \\ \vdots \\ a_{t-2} \\ a_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} a_t \\ y_t &= (-\theta_q, -\theta_{q-1}, \dots, -\theta_1) \mathbf{s}_t + a_t. \end{aligned} \quad (9.37)$$

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

Nest, consider the AR model, i.e.  $\phi(B)z_t = a_t$ . Aoki (1987) introduces two methods. The first method is a straightforward one by defining  $\mathbf{s}_t = (z_{t-p+1}, \dots, z_t)'$  to obtain

$$\begin{bmatrix} z_{t-p+2} \\ z_{t-p+3} \\ \vdots \\ z_{t+2} \\ z_{t+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & & 1 \\ \phi_p & \phi_{p-1} & \phi_{p-2} & \cdots & \phi_1 \end{bmatrix} \begin{bmatrix} z_{t-p+1} \\ z_{t-p+2} \\ \vdots \\ z_{t+1} \\ z_t \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} a_{t+1} \quad (9.38)$$

$$z_t = (0, 0, \dots, 0, 1)\mathbf{s}_t.$$

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, i.e.  $\mathbf{s}_t = z_t - a_t$  if  $p = 1$  and  $\mathbf{s}_t = (z_{t-p+1}, \dots, z_{t-1}, z_t - a_t)'$  if  $p > 1$ . Simple algebra shows that

$$\begin{bmatrix} z_{t-p+2} \\ z_{t-p+3} \\ \vdots \\ z_t \\ z_{t+1} - a_{t+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & & 1 \\ \phi_p & \phi_{p-1} & \phi_{p-2} & \cdots & \phi_1 \end{bmatrix} \begin{bmatrix} z_{t-p+1} \\ z_{t-p+2} \\ \vdots \\ z_{t-1} \\ z_t - a_t \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \phi_1 \end{bmatrix} a_t \quad (9.39)$$

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

Again,  $a_t$  appears in both transition and measurement equations.

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

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

Since  $z_t$  is an AR( $p$ ) model, we can use the transition equation in Eq. (9.38) or Eq. (9.39). If Eq. (9.38) is used, we can use  $y_t = \theta(B)z_t$  to construct the measurement equation as

$$y_t = (-\theta_{p-1}, -\theta_{p-2}, \dots, -\theta_1, 1)\mathbf{s}_t, \quad (9.40)$$

where it is understood that  $p > q$  and  $\theta_j = 0$  for  $j > q$ . On the other hand, if Eq. (9.39) is used as the transition equation, we construct the measurement equation as

$$y_t = (-\theta_{p-1}, -\theta_{p-2}, \dots, -\theta_1, 1)\mathbf{s}_t + a_t. \quad (9.41)$$

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. (9.23)-(9.24), 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.

## 9.4 Kalman Filter and Smoothing

In this section, we study Kalman filter and various smoothing methods for the general state-space model in Eq. (9.23)-(9.24). The derivation follows closely the steps taken in Section 9.1. 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).

### 9.4.1 Kalman filter

Recall that the object of Kalman filter is to obtain recursively the conditional distribution of  $\mathbf{s}_{t+1}$  given the data  $F_t = \{\mathbf{y}_1, \dots, \mathbf{y}_t\}$  and the model. Since all the distributions are normal, it suffices to study the conditional mean and conditional covariance matrix. Let  $\mathbf{s}_{j|i}$  and  $\Sigma_{j|i}$  be the conditional mean and covariance matrix of  $\mathbf{s}_j$  given  $F_i$ , i.e.  $\mathbf{s}_j|F_i \sim N(\mathbf{s}_{j|i}, \Sigma_{j|i})$ . From Eq. (9.23),

$$\mathbf{s}_{t+1|t} = E(\mathbf{d}_t + \mathbf{T}_t \mathbf{s}_t + \mathbf{R}_t \boldsymbol{\eta}_t | F_t) = \mathbf{d}_t + \mathbf{T}_t \mathbf{s}_{t|t}, \quad (9.42)$$

$$\Sigma_{t+1|t} = \text{Var}(\mathbf{T}_t \mathbf{s}_t + \mathbf{R}_t \boldsymbol{\eta}_t | F_t) = \mathbf{T}_t \Sigma_{t|t} \mathbf{T}_t' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t'. \quad (9.43)$$

Similarly to that of Section 9.1, let  $\mathbf{y}_{t|t-1}$  be the conditional mean of  $\mathbf{y}_t$  given  $F_{t-1}$ . From Eq. (9.24),

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

Let

$$\mathbf{v}_t = \mathbf{y}_t - \mathbf{y}_{t|t-1} = \mathbf{y}_t - (\mathbf{c}_t + \mathbf{Z}_t \mathbf{s}_{t|t-1}) = \mathbf{Z}_t (\mathbf{s}_t - \mathbf{s}_{t|t-1}) + \mathbf{e}_t, \quad (9.44)$$

be the one-step ahead forecast error of  $\mathbf{y}_t$  given  $F_{t-1}$ . It is easy to see that (a)  $E(\mathbf{v}_t | F_{t-1}) = \mathbf{0}$ , (b)  $\mathbf{v}_t$  is independent of  $F_{t-1}$ , i.e.  $\text{Cov}(\mathbf{v}_t, \mathbf{y}_j) = \mathbf{0}$  for  $1 \leq j < t$ , and (c)  $\{\mathbf{v}_t\}$  is a sequence of independent normal random vectors. Also, let  $\mathbf{V}_t = \text{Var}(\mathbf{v}_t)$  be the covariance matrix of the one-step ahead forecast error. From Eq. (9.44), we have

$$\mathbf{V}_t = \text{Var}[\mathbf{Z}_t (\mathbf{s}_t - \mathbf{s}_{t|t-1}) + \mathbf{e}_t] = \mathbf{Z}_t \Sigma_{t|t-1} \mathbf{Z}_t' + \mathbf{H}_t. \quad (9.45)$$

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

$$\begin{aligned} \mathbf{s}_{t|t} &= E(\mathbf{s}_t | F_t) = E(\mathbf{s}_t | F_{t-1}, \mathbf{v}_t) \\ &= E(\mathbf{s}_t | F_{t-1}) + \text{Cov}(\mathbf{s}_t, \mathbf{v}_t) [\text{Var}(\mathbf{v}_t)]^{-1} \mathbf{v}_t \\ &= \mathbf{s}_{t|t-1} + \mathbf{C}_t \mathbf{V}_t^{-1} \mathbf{v}_t, \end{aligned} \quad (9.46)$$

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

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

Here we assume that  $\mathbf{V}_t$  is invertible, because  $\mathbf{H}_t$  is. Using Eqs. (9.42) and (9.46), we obtain

$$\mathbf{s}_{t+1|t} = \mathbf{d}_t + \mathbf{T}_t \mathbf{s}_{t|t-1} + \mathbf{T}_t \mathbf{C}_t \mathbf{V}_t^{-1} \mathbf{v}_t = \mathbf{d}_t + \mathbf{T}_t \mathbf{s}_{t|t-1} + \mathbf{K}_t \mathbf{v}_t, \quad (9.47)$$

where

$$\mathbf{K}_t = \mathbf{T}_t \mathbf{C}_t \mathbf{V}_t^{-1} = \mathbf{T}_t \Sigma_{t|t-1} \mathbf{Z}_t' \mathbf{V}_t^{-1}, \quad (9.48)$$

which is the *Kalman gain* at time  $t$ .

Applying Theorem 1(2), we have

$$\begin{aligned}
\Sigma_{t|t} &= \text{Var}(\mathbf{s}_t|F_{t-1}) \\
&= \text{Var}(\mathbf{s}_t|F_{t-1}) - \text{Cov}(\mathbf{s}_t, \mathbf{v}_t)[\text{Var}(\mathbf{v}_t)]^{-1}\text{Cov}(\mathbf{s}_t, \mathbf{v}_t)' \\
&= \Sigma_{t|t-1} - \mathbf{C}_t \mathbf{V}_t^{-1} \mathbf{C}_t' \\
&= \Sigma_{t|t-1} - \Sigma_{t|t-1} \mathbf{Z}'_t \mathbf{V}_t^{-1} \mathbf{Z}_t \Sigma_{t|t-1}.
\end{aligned} \tag{9.49}$$

Plugging Eq. (9.49) in Eq. (9.43) and using Eq. (9.48) give

$$\Sigma_{t+1|t} = \mathbf{T}_t \Sigma_{t|t-1} \mathbf{L}'_t + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}'_t, \tag{9.50}$$

where

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

Putting the prior equations together, we obtain the celebrated Kalman filter for the state-space model in Eq. (9.23)-(9.24). Given the starting values  $\mathbf{s}_{1|0}$  and  $\Sigma_{1|0}$ , the Kalman filter algorithm is

$$\begin{aligned}
\mathbf{v}_t &= \mathbf{y}_t - \mathbf{c}_t - \mathbf{Z}_t \mathbf{s}_{t|t-1}, \\
\mathbf{V}_t &= \mathbf{Z}_t \Sigma_{t|t-1} \mathbf{Z}'_t + \mathbf{H}_t \\
\mathbf{K}_t &= \mathbf{T}_t \Sigma_{t|t-1} \mathbf{Z}'_t \mathbf{V}_t^{-1} \\
\mathbf{L}_t &= \mathbf{T}_t - \mathbf{K}_t \mathbf{Z}_t \\
\mathbf{s}_{t+1|t} &= \mathbf{d}_t + \mathbf{T}_t \mathbf{s}_{t|t-1} + \mathbf{K}_t \mathbf{v}_t, \\
\Sigma_{t+1|t} &= \mathbf{T}_t \Sigma_{t|t-1} \mathbf{L}'_t + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}'_t, \quad t = 1, \dots, T.
\end{aligned} \tag{9.51}$$

If the filtered quantities  $\mathbf{s}_{t|t}$  and  $\Sigma_{t|t}$  are also of interest, then we modify the filter to include the contemporaneous filtering equations in Eqs. (9.46) and (9.49). The resulting algorithm is

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

### ***Steady State***

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

$$\Sigma_* = \mathbf{T} \Sigma_* \mathbf{T}' - \mathbf{T} \Sigma_* \mathbf{Z} \mathbf{V}^{-1} \mathbf{Z}' \Sigma_* \mathbf{T}' + \mathbf{R} \mathbf{Q} \mathbf{R}'.$$

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

### 9.4.2 State estimation error and forecast error

Define the state prediction error as

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

From this definition, the covariance matrix of  $\mathbf{x}_t$  is  $\text{Var}(\mathbf{x}_t|F_{t-1}) = \text{Var}(\mathbf{s}_t|F_{t-1}) = \boldsymbol{\Sigma}_{t|t-1}$ . Similarly to Section 9.1, we investigate properties of  $\mathbf{x}_t$ . First, from Eq. (9.44),

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

Second, from Eqs. (9.51) and (9.23), and the prior equation, we have

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

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

$$\mathbf{v}_t = \mathbf{Z}_t\mathbf{x}_t + \mathbf{e}_t, \quad \mathbf{x}_{t+1} = \mathbf{L}_t\mathbf{x}_t + \mathbf{R}_t\boldsymbol{\eta}_t - \mathbf{K}_t\mathbf{e}_t, \quad (9.52)$$

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

Finally, similarly to the local-trend model in Section 9.1, we can show that the one-step forecast errors  $\{\mathbf{v}_t\}$  are independent of each other and  $\{\mathbf{v}_t, \dots, \mathbf{v}_T\}$  are independent of  $F_{t-1}$ .

### 9.4.3 State smoothing

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

$$\begin{aligned} \mathbf{s}_{t|T} &= E(\mathbf{s}_t|F_T) = E(\mathbf{s}_t|F_{t-1}, \mathbf{v}_t, \dots, \mathbf{v}_T) \\ &= E(\mathbf{s}_t|F_{t-1}) + \sum_{j=t}^T \text{Cov}(\mathbf{s}_t, \mathbf{v}_j)[\text{Var}(\mathbf{v}_t)]^{-1}\mathbf{v}_t \\ &= \mathbf{s}_{t|t-1} + \sum_{j=t}^T \text{Cov}(\mathbf{s}_t, \mathbf{v}_j)\mathbf{V}_t^{-1}\mathbf{v}_t, \end{aligned} \quad (9.53)$$

where the covariance matrices are conditional on  $F_{t-1}$ . The covariance matrices  $\text{Cov}(\mathbf{s}_t, \mathbf{v}_j)$  for  $j = t, \dots, T$  can be derived as follows. By Eq. (9.52)

$$\begin{aligned} \text{Cov}(\mathbf{s}_t, \mathbf{v}_j) &= E(\mathbf{s}_t\mathbf{v}_j') \\ &= E[\mathbf{s}_t(\mathbf{Z}_j\mathbf{x}_j + \mathbf{e}_j)'] = E(\mathbf{s}_t\mathbf{x}_j')\mathbf{Z}_j', \quad j = t, \dots, T. \end{aligned} \quad (9.54)$$

Furthermore,

$$\begin{aligned} E(\mathbf{s}_t\mathbf{x}_t') &= E[\mathbf{s}_t(\mathbf{s}_t - \mathbf{s}_{t|t-1})'] = \text{Var}(\mathbf{s}_t) = \boldsymbol{\Sigma}_{t|t-1}, \\ E(\mathbf{s}_t\mathbf{x}_{t+1}') &= E[\mathbf{s}_t(\mathbf{L}_t\mathbf{x}_t + \mathbf{R}_t\boldsymbol{\eta}_t - \mathbf{K}_t\mathbf{e}_t)'] = \boldsymbol{\Sigma}_{t|t-1}\mathbf{L}_t', \\ E(\mathbf{s}_t\mathbf{x}_{t+2}') &= \boldsymbol{\Sigma}_{t|t-1}\mathbf{L}_t'\mathbf{L}_{t+1}', \\ &\vdots = \vdots \\ E(\mathbf{s}_t\mathbf{x}_T') &= \boldsymbol{\Sigma}_{t|t-1}\mathbf{L}_t'\cdots\mathbf{L}_{T-1}'. \end{aligned} \quad (9.55)$$

Plugging the prior two equations in Eq. (9.53), we have

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

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

$$\mathbf{s}_{t|T} = \mathbf{s}_{t|t-1} + \boldsymbol{\Sigma}_{t|t-1} \mathbf{q}_{t-1}, \quad (9.56)$$

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

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

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

$$\mathbf{q}_{t-1} = \mathbf{Z}'_t \mathbf{V}_t^{-1} \mathbf{v}_t + \mathbf{L}'_t \mathbf{q}_t, \quad t = T, \dots, 1, \quad (9.57)$$

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

$$\mathbf{q}_{t-1} = \mathbf{Z}'_t \mathbf{V}_t^{-1} \mathbf{v}_t + \mathbf{L}'_t \mathbf{q}_t, \quad \mathbf{s}_{t|T} = \mathbf{s}_{t|t-1} + \boldsymbol{\Sigma}_{t|t-1} \mathbf{q}_{t-1}, \quad t = T, \dots, 1, \quad (9.58)$$

starting with  $\mathbf{q}_T = \mathbf{0}$ , where  $\mathbf{s}_{t|t-1}$ ,  $\boldsymbol{\Sigma}_{t|t-1}$ ,  $\mathbf{L}_t$ , and  $\mathbf{V}_t$  are available from the Kalman filter. This algorithm is referred to as the *fixed interval smoother* in the literature, e.g. 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 1(4) to the conditional joint distribution of  $\mathbf{s}_t$  and  $\{\mathbf{v}_t, \dots, \mathbf{v}_T\}$  given  $\mathbf{F}_{t-1}$ , we have

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

Using the covariance matrices in Eqs. (9.54)-(9.55), we further obtain

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

where

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

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

$$\mathbf{M}_{t-1} = \mathbf{Z}'_t \mathbf{V}_t^{-1} \mathbf{Z}_t + \mathbf{L}'_t \mathbf{M}_t \mathbf{L}_t, \quad t = T, \dots, 1, \quad (9.59)$$

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

$$\mathbf{M}_{t-1} = \mathbf{Z}'_t \mathbf{V}_t^{-1} \mathbf{Z}_t + \mathbf{L}'_t \mathbf{M}_t \mathbf{L}_t, \quad \Sigma_{t|T} = \Sigma_{t|t-1} \mathbf{M}_{t-1} \Sigma_{t|t-1}, \quad (9.60)$$

for  $t = T, \dots, 1$  with  $\mathbf{M}_T = \mathbf{0}$ . Note that, similarly to the local trend model in Section 9.1,  $\mathbf{M}_t = \text{Var}(\mathbf{q}_t)$ .

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

$$\begin{aligned} \mathbf{q}_{t-1} &= \mathbf{Z}'_t \mathbf{V}_t^{-1} \mathbf{v}_t + \mathbf{L}'_t \mathbf{q}_t, \\ \mathbf{s}_{t|T} &= \mathbf{s}_{t|t-1} + \Sigma_{t|t-1} \mathbf{q}_{t-1}, \\ \mathbf{M}_{t-1} &= \mathbf{Z}'_t \mathbf{V}_t^{-1} \mathbf{Z}_t + \mathbf{L}'_t \mathbf{M}_t \mathbf{L}_t \\ \Sigma_{t|T} &= \Sigma_{t|t-1} - \Sigma_{t|t-1} \mathbf{M}_{t-1} \Sigma_{t|t-1}, \quad t = T, \dots, 1, \end{aligned} \quad (9.61)$$

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

Suppose that the state-space model in Eqs. (9.23)-(9.24) is known. The application of Kalman filter and state smoothing proceeds in two steps. First, Kalman filter in Eq. (9.51) is used for  $t = 1, \dots, T$  and the quantities  $\mathbf{v}_t, \mathbf{V}_t, \mathbf{K}_t, \mathbf{s}_{t|t-1}, \Sigma_{t|t-1}$  are stored. Second, the state smoothing algorithm in Eq. (9.61) is applied for  $t = T, T-1, \dots, 1$  to obtain  $\mathbf{s}_{t|T}$  and  $\Sigma_{t|T}$ .

## 9.5 Forecasting

Suppose that the forecast origin is  $t$  and we are interested in predicting  $\mathbf{y}_{t+j}$  for  $j = 1, \dots, h$ , where  $h > 0$ . Also, we adopt the minimum mean square error forecasts. Similarly to the ARMA models, the  $j$ -step ahead forecast  $\mathbf{y}_t(j)$  turns out to be the expected value of  $\mathbf{y}_{t+j}$  given  $F_t$  and the model. That is,  $\mathbf{y}_t(j) = E(\mathbf{y}_{t+j}|F_t)$ . 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. (9.51) by treating  $\{\mathbf{y}_{t+1}, \dots, \mathbf{y}_{t+h}\}$  as missing values.

Consider the 1-step ahead forecast. From Eq. (9.24),

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

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

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

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

$$\text{Var}[\mathbf{e}_t(1)] = \mathbf{Z}_{t+1} \Sigma_{t+1|t} \mathbf{Z}'_{t+1} + \mathbf{H}_{t+1},$$

This is precisely the covariance matrix  $\mathbf{V}_{t+1}$  of the Kalman filter in Eq. (9.51). Thus, we have showed the case for  $h = 1$ .

Now, for  $h > 1$ , we consider 1-step to  $h$ -step forecasts sequentially. From Eq. (9.24), the  $j$ -step ahead forecast is

$$\mathbf{y}_t(j) = \mathbf{c}_{t+j} + \mathbf{Z}_{t+j} \mathbf{s}_{t+j|t}, \quad (9.62)$$

and the associated forecast error is

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

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

$$\text{Var}[\mathbf{e}_t(j)] = \mathbf{Z}_{t+j}\boldsymbol{\Sigma}_{t+j|t}\mathbf{Z}'_{t+j} + \mathbf{H}_{t+j}. \quad (9.63)$$

Furthermore, from Eq. (9.23),

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

which in turn implies that

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

Consequently,

$$\boldsymbol{\Sigma}_{t+j+1|t} = \mathbf{T}_{t+j}\boldsymbol{\Sigma}_{t+j|t}\mathbf{T}'_{t+j} + \mathbf{R}_{t+j}\mathbf{Q}_{t+j}\mathbf{R}'_{t+j}. \quad (9.64)$$

Noting that  $\text{Var}[\mathbf{e}_t(j)] = \mathbf{V}_{t+j}$ , Eqs. (9.62)-(9.64) are the recursion of the Kalman filter in Eq. (9.51) for  $t+j$  with  $j = 1, \dots, h$  when  $\mathbf{v}_{t+j} = \mathbf{0}$  and  $\mathbf{K}_{t+j} = \mathbf{0}$ . Thus, the forecast  $\mathbf{y}_t(j)$  and the covariance matrix of its forecast error  $\mathbf{e}_t(j)$  can be obtained via Kalman filter with missing values. Finally, the prediction error series  $\{\mathbf{v}_t\}$  can be used to evaluate the likelihood function for estimation and the standardized prediction errors  $\mathbf{D}_t^{-1}\mathbf{v}_t$  can be used for model checking, where  $\mathbf{D}_t = \text{diag}\{\mathbf{V}_t(1,1), \dots, \mathbf{V}_t(k,k)\}$  with  $\mathbf{V}_t(i,i)$  being the  $(i,i)$ th element of  $\mathbf{V}_t$ .

## 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-Verlag: New York.
- Anderson, B. D. O. and Moore, J. B. (1979), *Optimal Filtering*, Prentice-Hall: Englewood Cliffs.
- Chan, N. H. (2002), *Time Series: Applications to Finance*. John Wiley: New York.
- 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.
- Hamilton, J. (1994), *Time Series Analysis*. Princeton University Press: Princeton.
- Harvey, A. C. (1993), *Time Series Models*, 2nd edition. Harvester Wheatsheaf: Hemel Hempstead.
- 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 Press: New York.
- Kitagawa, G. and Gersch, W. (1996), *Smoothness Priors Analysis of Time Series*, Springer-Verlag: 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, <http://www.ssfpack.com/>.
- Shumway, R. H. and Stoffer, D. S. (2000), *Time Series Analysis and its Applications*. Springer-Verlag: New York.
- Tsay, R. S. (2005). *Analysis of Financial Time Series*, 2nd Edition. Wiley, Hoboken, NJ.
- West, M. and Harrison, J. (1997), *Bayesian Forecasting and Dynamic Models*, 2nd edition. Springer-Verlag: New York.