Chapter 10

State Space Models

10.1 State-Space representations

10.1.1 Introduction

A *state space model* for a (possibly multivariate) time series $\{\underline{Y}_t : t \in \mathbb{Z}\}$ consists of two equations. The first, known as the *observation equation* expresses the w-variate observation $\underline{Y}_t$ as a linear function of a v-variate state variable $\underline{X}_t$ plus noise. The second equation, the *state equation* determines the state $\underline{X}_{t+1}$ of the state variable in terms of the previous state $\underline{X}_t$ plus a noise term. Thus, a linear time homogeneous state space model is of the form:

$$\begin{cases} \underline{Y}_t = F_t \underline{X}_t + \underline{W}_t & t \in \mathbb{N} \quad \{\underline{W}_t\} \sim \text{WN}(\underline{0}, \{R_t\}) \\ \underline{X}_{t+1} = G_t \underline{X}_t + \underline{V}_{t+1} & t \in \mathbb{N} \quad \{\underline{V}_t\} \sim \text{WN}(\underline{0}, \{Q_t\}). \end{cases} \tag{10.1}$$

where $\{\underline{Y}_t\}$ is the observation process, $\{\underline{X}_t\}$ is a v-variate process describing the state of an underlying system, $\{F_t\}$ is a sequence of $v \times v$ matrices and $\{G_t\}$ is a sequence of $w \times v$ matrices. The state $\{\underline{X}_t\}$ cannot be observed directly.

We consider the setting where the 'noise' $\{\underline{W}_t\}$ and the 'noise' $\{\underline{V}_t\}$ may be correlated;

$$\left(\frac{W_t}{V_t}\right) \sim WN\left(\left(\frac{0}{0}\right), \left(\begin{array}{c|c} R_t & S_t \\ \hline S_t' & Q_t \end{array}\right)\right). \tag{10.2}$$

To complete the specification it is assumed that the initial state $\underline{X}_1$ is uncorrelated with $\{\underline{W}_t\}$ and $\{\underline{V}_t\}$.

10.1.2 Examples of State Space Representations

We now show that processes we have already encountered have a state space representation.

Example 10.1 (State space representation of an AR(1) process)**.**

A simple example of a Time Series model is the causal AR(1) process. Let Y be a stationary process satisfing:

$$Y_t = \phi Y_{t-1} + Z_t, \quad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2). \tag{10.3}$$

for $|\phi| < 1$. This process already has a state space representation, since $\{Y_t\}$ is already of the form of the second equation of (10.1). The state space representation is (trivially):

$$\begin{cases} Y_t = X_t \\ X_{t+1} = \phi X_t + Z_{t+1} \qquad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2). \end{cases} \tag{10.4}$$

Matching the components, this corresponds to: $G_t = 1$ for all t and $R = 0$ for the observation equation of (10.1) and $F_t = \phi$ for all t and $Q = \sigma^2$ for the state equation of (10.1), so $V_t = Z_t$.

It is assumed that the Kalman recursion starts from time 0, so set $X_1 = Y_1 = \sum_{j=0}^{\infty} \phi^j Z_{1-j}$. Then Equation (10.1) becomes (10.4) and it is clear that $\{Y_t : t \in \mathbb{N}\}$ satisfies Equation (10.3). □

Example 10.2 (State space representation of an AR(p) process)**.**

Now consider a causal AR(p) process:

$$Y_t = \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + Z_t, \quad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2).$$

With $\underline{X}_t$ thus defined, the *state* which determines the observation $\{Y_t\}$ is the whole vector $\underline{X}_t = (Y_{t-p+1}, \ldots, Y_t)^t$ and the state equation in (10.3) gives a recursive relation for this vector. The observation equation is (trivially)

$$Y_t = (0, \ldots, 0, 1)\underline{X}_t.$$

and the state equation for $\underline{X}_{t+1}$ may be constructed quite easily as:

$$\underline{X}_{t+1} = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \ldots & 1 \\ \phi_p & \phi_{p-1} & \phi_{p-2} & \ldots & \phi_1 \end{pmatrix} \underline{X}_t + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} Z_{t+1}. \tag{10.5}$$

These equations have the required form with $\underline{W}_t = \underline{0}$ and

$$\underline{V}_{t+1} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ Z_{t+1} \end{pmatrix}.$$

It only remains to specify $\underline{X}_1$, which follows in the same way as for $p = 1$.

Clearly, the initial condition is unnecessary if the state space representation is taken for $t \in \mathbb{Z}$. □

Example 10.3 (State space representation for ARMA(p,q) process)**.**

Let $\{Y_t\}$ be a causal ARMA(p,q) process satisfying

$$\phi(B)Y_t = \theta(B)Z_t, \quad \{Z_t\} \sim \text{WN}(0, \sigma^2).$$

In the following state space representation, the AR part is incorporated into the state equation, as in Example 10.2 for the AR(p) process, while the MA part is incorporated into the observation equation.

This is obtained by making the following observation: let $U_t = \phi^{-1}(B)Z_t$ (this is well defined for ϕ causal). Then U_t is an AR(p) process satisfying $\phi(B)U_t = Z_t$ and $Y_t = \theta(B)U_t$. This follows from:

$$\phi(B)Y_t = \theta(B)Z_t = \theta(B)\phi(B)U_t = \phi(B)\theta(B)U_t$$

so that, applying $\phi(B)^{-1}$ to both sides:

$$Y_t = \theta(B)U_t.$$

Therefore, a state space representation can be derived from:

$$\begin{cases} Y_t = \theta(B)U_t \\ \phi(B)U_{t+1} = Z_{t+1} \Rightarrow U_{t+1} = \sum_{j=1}^{p} \phi_j U_{t+1-j} + Z_{t+1}. \end{cases}$$

The state vector is $\underline{U}_t$ and, rewriting the first of these equations:

$$Y_t = (\theta_q, \theta_{q-1}, \dots, \theta_1, 1) \begin{pmatrix} U_{t-q} \\ \vdots \\ U_t \end{pmatrix}.$$

Let $r = \max(p, q+1)$. The state vector is $X_t = (U_{t-r+1}, \dots, U_t)^t$. If $r = p$, then the state equation is simply Equation (10.5). If $q + 1 > p$, then the state equation is:

$$\underline{X}_{t+1} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 & \phi_p & \phi_{p-1} & \phi_{p-2} & \dots & \phi_1 \end{pmatrix} \underline{X}_t + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} Z_{t+1}$$

and the observation equation:

$$Y_t = (1, \theta_1, \ldots, \theta_q)\underline{X}_t.$$

□

10.1.3 Akaike's State Space Model

This is a state space model for producing forecasts from an ARMA process. Akaike defines the state vector X_t as the minimum collection of variables that contain all the information needed to produce forecasts at the forecast origin t (i.e. if $\{Y_t\}$ is the ARMA process, then X_t contains all $(Y_t, Y_{t-1}, Y_{t-2}, \ldots)$ necessary for forecasts $\widehat{Y}_{t+h|t} = \mathbb{E}[Y_{t+h}|\mathcal{F}_t^{(Y)}]$). For the ARMA process with $r = \max(p, q+1)$,

$$X_t = (\widehat{Y}_{t|t}, \widehat{Y}_{t+1|t}, \ldots, \widehat{Y}_{t+r-1|t})'$$

where x' denotes the transpose of vector x.

Note: $\widehat{Y}_{t|t} = Y_t$. Therefore, with this state vector, the observation equation is:

$$Y_t = vX_t \qquad v = (1, 0, \ldots, 0).$$

We use the notation $X_{j,t}$ to denote the jth component of state vector X_t. From the definition,

$$X_{1,t+1} = Y_{t+1} = \widehat{Y}_{t+1|t} + (Y_{t+1} - \widehat{Y}_{t+1|t}) = X_{2,t} + Z_{t+1}.$$

This follows, because

$$Y_{t+1} = \sum_{j=1}^{p} \phi_j Y_{t+1-j} + Z_{t+1} + \sum_{j=1}^{q} \theta_j Z_{t+1-j},$$

hence

$$\widehat{Y}_{t+1|t} = \mathbb{E}[Y_{t+1}|\mathcal{F}_t] = \sum_{j=1}^{p} \phi_j Y_{t+1-j} \sum_{j=1}^{q} \theta_j Z_{t+1-j}$$

and the result follows by subtraction.

Now consider the expansion of a causal ARMA process:

$$Y_t = \sum_{i=0}^{\infty} \psi_i Z_{t-i} \tag{10.6}$$

where $\psi_0 = 1$ and the other weights can be computed from $\psi(z) = \frac{\theta(z)}{\phi(z)}$. This follows from expanding

$$1 + \sum_{j=1}^{\infty} \psi_j z^j = \frac{1 + \sum_{j=1}^{q} \theta_j z^j}{1 - \sum_{j=1}^{p} \phi_j z^j} = 1 + z(\theta_1 + \phi_1) + z^2(\theta_2 + \phi_2 + \theta_1\phi_1) + \ldots$$

so that

$$\psi_1 = \phi_1 + \theta_1, \quad \psi_2 = \phi_1\psi_1 + \phi_2 + \theta_2$$

and recursively

$$\psi_{r-1} = \sum_{i=1}^{r-1} \phi_i \psi_{r-1-i} + \theta_{r-1}.$$

Using expansion

$$\widehat{Y}_{t+j|t+1} = \mathbb{E}[Y_{t+j}|\mathcal{F}_{t+1}] = \sum_{i=j-1}^{\infty} \psi_i Z_{t+j-i} = \psi_{j-1} Z_{t+1} + \widehat{Y}_{t+j|t}.$$

Since

$$Y_t = \sum_{j=1}^{p} \phi_j Y_{t-j} + Z_t + \sum_{j=1}^{q} \theta_j Z_{t-j},$$

taking expectations gives:

$$\widehat{Y}_{t+r|t+1} = \sum_{i=1}^{r} \phi_i \widehat{Y}_{t+r-i|t+1} + \theta_{r-1} Z_{t+1}.$$

Recall that $r = p$ or $q+1$. $\phi_i = 0$ for $i \geq p+1$ and $\theta_i = 0$ for $i \geq q+1$. Inserting these gives:

$$\begin{aligned}
\widehat{Y}_{t+r|t+1} &= \sum_{i=1}^{r-1} \phi_i (\widehat{Y}_{t+r-i|t} + \psi_{r-i-1} Z_{t+1}) + \phi_r \widehat{Y}_{t|t} - \theta_{r-1} Z_{t+1} \\
&= \sum_{i=1}^{r} \phi_i \widehat{Y}_{t+r-i|t} + \left(\sum_{i=1}^{r-1} \phi_i \psi_{r-1-i} + \theta_{r-1} \right) Z_{t+1} \\
&= \sum_{i=1}^{r} \phi_i \widehat{Y}_{t+r-i|t} + \psi_{r-1} Z_{t+1}.
\end{aligned}$$

Therefore, the state vector X_t satisfies the following:

$$\begin{pmatrix} y_{t+1} \\ y_{t+2|t+1} \\ \vdots \\ y_{t+m|t+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ \phi_m & \phi_{m-1} & \phi_{m-2} & \dots & \phi_1 \end{pmatrix} \begin{pmatrix} y_t \\ y_{t+1|t} \\ \vdots \\ y_{t+m-1|t} \end{pmatrix} + \begin{pmatrix} 1 \\ \psi_1 \\ \vdots \\ \psi_{m-2} \\ \psi_{m-1} \end{pmatrix} Z_{t+1}$$

This gives a recursive method of computing the forecasts from forecast origin t as t evolves. □

10.1.4 Local Trend Model

Consider the univariate time series $\{Y_t : t \in \mathbb{Z}\}$ satisfying:

$$\begin{cases} y_t = \mu_t + e_t & e_t \sim N(0, \sigma_e^2) \\ \mu_{t+1} = \mu_t + \eta_t & \eta_t \sim N(0, \sigma_\eta^2) \end{cases}$$

where $\{e_t\}$ and $\{\eta_t\}$ are two independent Gaussian noise processes. This is related to the ARIMA model as follows: if there is no *measurement* error (i.e. $\sigma_e = 0$) then $y_t = \mu_t$ which is an ARIMA(0,1,0) model. If $\sigma_e > 0$, there is an extra measurement error. Consequently, y_t is an ARIMA(0,1,1) process satisfying:

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

$$(I - B)y_t = (1 - \theta B)a_t$$

where $\{a_t\}$ is a Gaussian white noise with variance σ_a^2 which may be computed from considering the covariance structure of an MA(1) process: if Z_t is $WN(0, \sigma^2)$ and $X_t = Z_t + \theta Z_{t-1}$ then

$$\text{Var}(X_t) = (1 + \theta^2)\sigma^2 \qquad \text{Cov}(X_{t-1}, X_t) = \theta\sigma^2.$$

It follows that:

$$2\sigma_e^2 + \sigma_\eta^2 = (1 + \theta^2)\sigma_a^2 \qquad \sigma_e^2 = \theta\sigma_a^2.$$

This is easily solvable; select the solution satisfying $|\theta| < 1$. This gives:

$$\sigma_a = \frac{\sigma_\eta}{2} + \sqrt{\sigma_e^2 + \frac{\sigma_\eta^2}{4}} \qquad \theta = 1 - \frac{\sigma_\eta}{\sigma_a} = 1 - \frac{2}{1 + \sqrt{1 + \frac{4\sigma_e^2}{\sigma_\eta^2}}}.$$

10.1.5 CAPM (Capital Asset Pricing Model) with Time-Varying Coefficients

We now show that the CAPM may be written in this form. The model is:

$$\begin{cases} r_t = \alpha_t + \beta_t r_{Mt} + e_t & e_t \sim IIDN(0, \sigma_e^2) \\ \alpha_{t+1} = \alpha_t + \eta_t & \eta_t \sim IIDN(0, \sigma_\eta^2) \\ \beta_{t+1} = \beta_t + \epsilon_t & \eta_t \sim IIDN(0, \sigma_\epsilon^2) \end{cases}$$

Here r_t is the *excess return* of an asset (this is the return that exceeds what was expected), while r_{Mt} is the excess return of the market, α is the risk-free return and β is the asset's price volatility with respect to the overall market.

Firstly, consider α and β;

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

The return r_t may be written as:

$$r_t = (1, r_{Mt}) \begin{pmatrix} \alpha_t \\ \beta_t \end{pmatrix} + e_t.$$

This is a state space model with $F = (1, r_{Mt})$, $\{W_t\} = \{e_t\} \sim IIDN(0, \sigma_e^2)$, $G = I_2$,

$$\{V_t\} = \left\{ \begin{pmatrix} \eta_t \\ \epsilon_t \end{pmatrix} \right\} \sim IIDN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_\epsilon^2 \end{pmatrix} \right)$$