Lecture 12: State Space Model and Kalman Filter

Bus 41910, Time Series Analysis, Mr. R. Tsay

1 Introduction

A state space model consists of two equations:

$$S_{t+1} = FS_t + Ge_t \quad (1)$$
$$Z_t = HS_t + \epsilon_t \quad (2)$$

where S_t is a state vector of dimension m, Z_t is the observed time series, F, G, H are matrices of parameters, $\{e_t\}$ and $\{\epsilon_t\}$ are *iid* random vectors satisfying

$$E(e_t) = 0, \quad E(\epsilon_t) = 0, \quad \mathrm{Cov}(e_t) = Q, \quad \mathrm{Cov}(\epsilon_t) = R$$

and $\{e_t\}$ and $\{\epsilon_t\}$ are independent. In the engineering literature, a state vector denotes the unobservable vector that describes the "status" of the system. Thus, a state vector can be thought of as a vector that contains the necessary information to predict the future observations (i.e., minimum mean squared error forecasts).

In this course, Z_t is a scalar and F, G, H are constants. A general state space model in fact allows for vector time series and time-varying parameters. Also, the independence requirement between e_t and ϵ_t can be relaxed so that e_t and ϵ_t are correlated.

2 Relation to ARMA models

To appreciate the above state space model, we first consider its relation with ARMA models. The basic relations are

- an ARMA model can be put into a state space form in "infinite" many ways;
- for a given state space model in (1)-(2), there is an ARMA model.

A. State space model to ARMA model:

The key here is the Cayley-Hamilton theorem, which says that for any $m \times m$ matrix F with characteristic equation

$$c(\lambda) = |F - \lambda I| = \lambda^m + \alpha_1\lambda^{m-1} + \alpha_2\lambda^{m-2} + \cdots + \alpha_{m-1}\lambda + \alpha_0,$$

we have $c(F) = 0$. In other words, the matrix F satisfies its own characteristic equation, i.e.

$$F^m + \alpha_1 F^{m-1} + \alpha_2 F^{m-2} + \cdots + \alpha_{m-1}F + \alpha_m I = 0.$$

Next, from the state transition equation, we have

$$\begin{aligned}
S_t &= S_t \\
S_{t+1} &= FS_t + Ge_t \\
S_{t+2} &= F^2S_t + FGe_t + Ge_{t+1} \\
S_{t+3} &= F^3S_t + F^2Ge_t + FGe_{t+1} + Ge_{t+2} \\
\vdots &= \vdots \\
S_{t+m} &= F^mS_t + F^{m-1}Ge_t + \cdots + FGe_{t+m-2} + Ge_{t+m-1}.
\end{aligned}$$

Multiplying the above equations by $\alpha_m, \alpha_{m-1}, \cdots, \alpha_1, 1$, respectively, and summing, we obtain

$$S_{t+m} + \alpha_1 S_{t+m-1} + \cdots + \alpha_{m-1}S_{t+1} + \alpha_m S_t = Ge_{t+m-1} + \beta_1 e_{t+m-2} + \cdots + \beta_{m-1}e_t. \quad (3)$$

In the above, we have used the fact that $c(F) = 0$.
Finally, two cases are possible. First, assume that there is no observational noise, i.e. $\epsilon_t = 0$ for all t in (2). Then, by multiplying H from the left to equation (3) and using $Z_t = HS_t$, we have

$$Z_{t+m} + \alpha_1 Z_{t+m-1} + \cdots + \alpha_{m-1}Z_{t+1} + \alpha_m Z_t = a_{t+m} - \theta_1 a_{t+m-1} - \cdots - \theta_{m-1}a_{t+1},$$

where $a_t = HGe_{t-1}$. This is an ARMA$(m, m-1)$ model.
The second possibility is that there is an observational noise. Then, the same argument gives

$$(1 + \alpha_1 B + \cdots + \alpha_m B^m)(Z_{t+m} - \epsilon_{t+m}) = (1 - \theta_1 B - \cdots - \theta_{m-1}B^{m-1})a_{t+m}.$$

By combining ϵ_t with a_t, the above equation is an ARMA(m, m) model.

B. ARMA model to state space model:
We begin the discussion with some simple examples. Three general approaches will be given later.
Example 1: Consider the AR(2) model

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t.$$

For such an AR(2) process, to compute the forecasts, we need Z_{t-1} and Z_{t-2}. Therefore, it is easily seen that

$$\begin{bmatrix} Z_{t+1} \\ Z_t \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} Z_t \\ Z_{t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} e_t,$$

where $e_t = a_{t+1}$ and

$$Z_t = [1, 0]S_t$$

where $S_t = (Z_t, Z_{t-1})'$ and there is no observational noise.

Example 2: Consider the MA(2) model

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}.$$

<u>Method 1</u>:

$$\begin{bmatrix} a_t \\ a_{t-1} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a_{t-1} \\ a_{t-2} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} a_t$$
$$Z_t = [-\theta_1, -\theta_2] S_t + a_t.$$

Here the innovation a_t shows up in both the state transition equation and the observation equation. The state vector is of dimension 2.
<u>Method 2</u>: For an MA(2) model, we have

$$\begin{aligned} Z_{t|t} &= Z_t \\ Z_{t+1|t} &= -\theta_1 a_t - \theta_2 a_{t-1} \\ Z_{t+2|t} &= -\theta_2 a_t. \end{aligned}$$

Let $S_t = (Z_t, -\theta_1 a_t - \theta_2 a_{t-1}, -\theta_2 a_t)'$. Then,

$$S_{t+1} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} S_t + \begin{bmatrix} 1 \\ -\theta_1 \\ -\theta_2 \end{bmatrix} a_{t+1}$$

and

$$Z_t = [1, 0, 0] S_t.$$

Here the state vector is of dimension 3, but there is no observational noise.

Exercise: Generalize the above result to an MA(q) model.

Next, we consider three general approaches.
Akaike's approach: For an ARMA(p, q) process, let $m = \max\{p, q+1\}$, $\phi_i = 0$ for $i > p$ and $\theta_j = 0$ for $j > q$. Define $S_t = (Z_t, Z_{t+1|t}, Z_{t+2|t}, \cdots, Z_{t+m-1|t})'$ where $Z_{t+\ell|t}$ is the conditional expectation of $Z_{t+\ell}$ given $\Psi_t = \{Z_t, Z_{t-1}, \cdots\}$. By using the updating equation of forecasts (recall what we discussed before)

$$Z_{t+1}(\ell - 1) = Z_t(\ell) + \psi_{\ell-1} a_{t+1},$$

it is easy to show that

$$S_{t+1} = F S_t + G a_{t+1}$$
$$Z_t = [1, 0, \cdots, 0] S_t$$

where

$$F = \left[\begin{array}{c|cccc} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & & \\ \phi_m & \phi_{m-1} & \cdots & \phi_2 & \phi_1 \end{array}\right], \quad G = \begin{bmatrix} 1 \\ \psi_1 \\ \psi_2 \\ \vdots \\ \psi_{m-1} \end{bmatrix}.$$

The matrix F is call a companion matrix of the polynomial $1 - \phi_1 B - \cdots - \phi_m B^m$.

Aoki's Method: This is a two-step procedure. First, consider the MA(q) part. Letting $W_t = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$, we have

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

$$W_t = [-\theta_1, -\theta_2, \cdots, -\theta_q] S_t + a_t.$$

In the next step, we use the usual AR(p) format for

$$Z_t - \phi_1 Z_{t-1} - \cdots - \phi_p Z_{t-p} = W_t.$$

Consequently, define the state vector as

$$S_t = (Z_{t-1}, Z_{t-2}, \cdots, Z_{t-p}, a_{t-1}, \cdots, a_{t-q})'.$$

Then, we have

$$\begin{bmatrix} Z_t \\ Z_{t-1} \\ \vdots \\ Z_{t-p+1} \\ a_t \\ a_{t-1} \\ \vdots \\ a_{t-q+1} \end{bmatrix} = \left[\begin{array}{cccc|cccc} \phi_1 & \phi_2 & \cdots & \phi_p & -\theta_1 & -\theta_1 & \cdots & -\theta_q \\ 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & & & & \vdots & & & \\ 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & 0 & & & \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 1 & 0 \end{array}\right] \begin{bmatrix} Z_{t-1} \\ Z_{t-2} \\ \vdots \\ Z_{t-p} \\ a_{t-1} \\ a_{t-2} \\ \vdots \\ a_{t-q} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} a_t.$$

and

$$Z_t = [\phi_1, \cdots, \phi_p, -\theta_1, \cdots, -\theta_q] S_t + a_t.$$

Third approach: The third method is used by some authors, e.g. Harvey and his associates. Consider an ARMA(p, q) model

$$Z_t = \sum_{i=1}^{p} \phi_i Z_{t-i} + a_t - \sum_{j=1}^{q} \theta_j a_{t-j}.$$

Let $m = \max\{p, q\}$. Define $\phi_i = 0$ for $i > p$ and $\theta_j = 0$ for $j > q$. The model can be written as

$$Z_t = \sum_{i=1}^{m} \phi_i Z_{t-i} + a_t - \sum_{i=1}^{m} \theta_i a_{t-i}.$$

Using $\psi(B) = \theta(B)/\phi(B)$, we can obtain the ψ-weights of the model by equating coefficients of B^j in the equation

$$(1 - \theta_1 B - \cdots - \theta_m B^m) = (1 - \phi_1 B - \cdots - \phi_m B^m)(\psi_0 + \psi_1 B + \cdots + \psi_m B^m + \cdots),$$

where $\psi_0 = 1$. In particular, consider the coefficient of B^m, we have

$$-\theta_m = -\phi_m \psi_0 - \phi_{m-1}\psi_1 - \cdots - \phi_1 \psi_{m-1} + \psi_m.$$

Consequently,

$$\psi_m = \sum_{i=1}^{m} \phi_i \psi_{m-i} - \theta_m. \tag{4}$$

Next, from the ψ-weight representation

$$Z_{t+m-i} = a_{t+m-i} + \psi_1 a_{t+m-i-1} + \psi_2 a_{t+m-i-2} + \cdots,$$

we obtain

$$\begin{aligned} Z_{t+m-i|t} &= \psi_{m-i} a_t + \psi_{m-i+1} a_{t-1} + \psi_{m-i+2} a_{t-2} + \cdots \\ Z_{t+m-i|t-1} &= \psi_{m-i+1} a_{t-1} + \psi_{m-i+2} a_{t-2} + \cdots. \end{aligned}$$

Consequently,

$$Z_{t+m-i|t} = Z_{t+m-i|t-1} + \psi_{m-i} a_t, \quad m - i > 0. \tag{5}$$

We are ready to setup a state space model. Define $S_t = (Z_{t|t-1}, Z_{t+1|t-1}, \cdots, Z_{t+m-1|t-1})'$. Using $Z_t = Z_{t|t-1} + a_t$, the observational equation is

$$Z_t = [1, 0, \cdots, 0] S_t + a_t.$$

The state-transition equation can be obtained by Equations (5) and (4). First, for the first $m - 1$ elements of S_{t+1}, Equation (5) applies. Second, for the last element of S_{t+1}, the model implies

$$Z_{t+m|t} = \sum_{i=1}^{m} \phi_i Z_{t+m-i|t} - \theta_m a_t.$$

Using Equation (5), we have

$$\begin{aligned} Z_{t+m|t} &= \sum_{i=1}^{m} \phi_i (Z_{t+m-i|t-1} + \psi_{m-i} a_t) - \theta_m a_t \\ &= \sum_{i=1}^{m} \phi_i Z_{t+m-i|t-1} + (\sum_{i=1}^{m} \phi_i \psi_{m-i} - \theta_m) a_t \\ &= \sum_{i=1}^{m} \phi_i Z_{t+m-i|t-1} + \psi_m a_t, \end{aligned}$$

where the last equality uses Equation (4). Consequently, the state-transition equation is

$$S_{t+1} = FS_t + Ga_t$$

where

$$F = \left[\begin{array}{c|ccccc} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & & \\ \phi_m & \phi_{m-1} & \cdots & \phi_2 & \phi_1 \end{array} \right], \quad G = \left[\begin{array}{c} \psi_1 \\ \psi_2 \\ \psi_3 \\ \vdots \\ \psi_m \end{array} \right].$$

Note that for this third state space model, the dimension of the state vector is $m = \max\{p, q\}$, which may be lower than that of the Akaike's approach. However, the innovations to both the state-transition and observational equations are a_t.

3 Kalman Filter

Kalman filter is a set of recursive equations that provide a simple way to update the information in a state space model. It basically decomposes an observation into conditional mean and predictive residual sequentially. [Cholesky decomposition.] Thus, it has wide applications in statistical analysis.
The simplest way to derive the Kalman recursion is to use normality assumption. It should be pointed out, however, that the recursion is a result of the least squares principle (or projection) not normality. Thus, the recursion continues to hold for non-normal case. The only difference is that the solution obtained is only optimal within the class of linear solutions. With normailty, the solution is optimal among all possible solutions (linear and nonlinear).
Under normality, we have

- that normal prior plus normal likelihood results in a normal posterior,
- that if the random vector (X, Y) are jointly normal

  $$\left[\begin{array}{c} X \\ Y \end{array} \right] \sim N(\left[\begin{array}{c} \mu_x \\ \mu_y \end{array} \right], \left[\begin{array}{cc} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{array} \right]),$$

  then the conditional distribution of X given $Y = y$ is normal

  $$X|Y = y \sim N[\mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y), \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}].$$

Using these two results, we are ready to derive the Kalman filter. In what follows, let $P_{t+j|t}$ be the conditional covariance matrix of S_{t+j} given $\{Z_t, Z_{t-1}, \cdots\}$ for $j \geq 0$ and $S_{t+j|t}$ be the conditional mean of S_{t+j} given $\{Z_t, Z_{t-1}, \cdots\}$.

First, by the state space model, we have

$$\begin{aligned}
S_{t+1|t} &= FS_{t|t} && (6)\\
Z_{t+1|t} &= HS_{t+1|t} && (7)\\
P_{t+1|t} &= FP_{t|t}F' + GQG' && (8)\\
V_{t+1|t} &= HP_{t+1|t}H' + R && (9)\\
C_{t+1|t} &= HP_{t+1|t} && (10)
\end{aligned}$$

where $V_{t+1|t}$ is the conditional variance of Z_{t+1} given $\{Z_t, Z_{t-1}, \cdots\}$ and $C_{t+1|t}$ denotes the conditional covariance between Z_{t+1} and S_{t+1}. Next, consider the joint conditional distribution between S_{t+1} and Z_{t+1}. The above results give

$$\begin{bmatrix} S_{t+1} \\ Z_{t+1} \end{bmatrix}_t \sim N(\begin{bmatrix} S_{t+1|t} \\ Z_{t+1|t} \end{bmatrix}, \begin{bmatrix} P_{t+1|t} & P_{t+1|t}H' \\ HP_{t+1|t} & HP_{t+1|t}H' + R \end{bmatrix}).$$

Finally, when Z_{t+1} becomes available, we may use the property of nromality to update the distribution of S_{t+1}. More specifically,

$$S_{t+1|t+1} = S_{t+1|t} + P_{t+1|t}H'[HP_{t+1|t}H' + R]^{-1}(Z_{t+1} - Z_{t+1|t}) \tag{11}$$

and

$$P_{t+1|t+1} = P_{t+1|t} - P_{t+1|t}H'[HP_{t+1|t}H' + R]^{-1}HP_{t+1|t}. \tag{12}$$

Obviously,

$$r_{t+1|t} = Z_{t+1} - Z_{t+1|t} = Z_{t+1} - HS_{t+1|t}$$

is the predictive residual for time point $t+1$. The updating equation in (11) says that when the predictive residual $r_{t+1|t}$ is non-zero there is new information about the system so that the state vector should be modified. The contribution of $r_{t+1|t}$ to the state vector, of course, needs to be weighted by the variance of $r_{t+1|t}$ and the conditional covariance matrix of S_{t+1}.

In summary, the Kalman filter consists of the following equations:

- Prediction: (6), (7), (8) and (9)
- Updating: (11) and (12).

In practice, one starts with initial prior information $S_{0|0}$ and $P_{0|0}$, then predicts $Z_{1|0}$ and $V_{1|0}$. Once the observation Z_1 is available, uses the updating equations to compute $S_{1|1}$ and $P_{1|1}$, which in turns serve as prior for the next observation. This is the Kalman recusion.

Specifically, let $S_{0|0}$ and $P_{0|0}$ be some arbitrary initial values. From Eqs. (6) and (7), we obtain the predictions $S_{1|0}$ and $Z_{1|0}$. From Eq. (8), we obtain $P_{1|0}$, which in turn via Eqs. (9) and (10), we get $V_{1|0}$ and $C_{1|0}$. Now, suppose Z_1 is observed. We can compute the residual $r_{1|0} = Z_1 - Z_{1|0}$. Using this residual and Eqs. (11) and (12), we can update the state vector, i.e. we have $S_{1|1}$ and $P_{1|1}$. This completes one iteration of Kalman filter.

Note that the effect of the initial values $S_{0|0}$ and $P_{0|0}$ is decresing as t increases. The main reason is that for a stationary time series, all eigenvalues of the coefficient matrix F are less than one in modulus. As such, Kalman filter recursion ensures that the effect of the initial values indeed vanishes as t increases.

4 Statistical Inference

A key objective of State-space models is to infer properties of the state S_t based on the data $\{z_1, \ldots, z_t\}$ and the model. Three types of inference are commonly discussed in the literature. They are *filtering, prediction* and *smoothing*. Let $F_t = \sigma - \{z_t, z_{t-1}, \ldots\}$ be the information available at time t (inclusive), and assume that the model is known (i.e. all parameters of the model are known). The three types of inference are

- Filtering: To recover the state vector S_t given F_t,
- Prediction: To predict S_{t+h} or z_{t+h} for $h > 0$, given F_t,
- Smoothing: To estimate S_t given F_T, where $T > t$.

We look at a simple example in detail to understand these three types of inference. See Chapter 11 of Tsay (2005).
Based on the discussion of Kalman filter, the applications of filtering and prediction are easy to understand. The smoothing, on the other hand, requires some further explanation. First, consider the predictive residual $r_t = Z_t - Z_{t|t-1} = Z_t - HS_{t|t-1}$. It is easy to see that $r_t = H(S_t - S_{t|t-1}) + \epsilon_t$ and the conditional variance of Z_t given F_{t-1} is $V_{t|t-1} = HP_{t|t-1}H' + R$. (See Eq. (9). Also, we have $E(r_t) = 0$ and $\text{Cov}(r_t, Z_j) = E(r_t Z_j) = E[E(r_t Z_j|F_{t-1})] = E[Z_j E(r_t|F_{t-1})] = 0$ for all $j < t$. In other words, as expected, the predictive residual r_t is uncorrleated with the past observations Z_j for $j < t$.
In fact, the series of predictive residuals $\{r_1, r_2, \ldots, r_T\}$, where T is the sample size, has some nice properties:

1. $\{r_1, \ldots, r_T\}$ are serially uncorrelated (independent under normality) and they are linear function of $\{Z_1, Z_2, \ldots, Z_T\}$.
2. The transformation from $\{Z_1, \ldots, Z_T\}$ to $\{r_1, \ldots, r_T\}$ has unity Jacobian. Consequently, $p(Z_1, \ldots, Z_T) = p(r_1, \ldots, r_T) = \prod_{t=1}^{T} p(r_t)$. This provides a simple way to evaluate the likelihood function of the data.
3. The σ-field $F_T = \sigma\{Z_1, \ldots, Z_T\} = \sigma\{Z_1, \ldots, Z_{t-1}, r_t, \ldots, r_T\}$.

Recall that $S_{t|j}$ and $P_{t|j}$ are the conditional expectation and covariance matrix of S_t given F_j. The smoothing is concerned with $S_{t|T}$ and $P_{t|T}$. It turns out that these two quantities can be obtained recursively from the results of Kalman filter, i.e., from $S_{t|t-1}$ and $P_{t|t-1}$. The recursion is similar to the Kalman filter we discussed so far. See Section 11.4 of Tsay (2005) for details.