Chapter 11

Kalman Recursions

State space models can simplify some problems, such as maximum-likelihood estimation and handling missing values. Three problems for estimation of $\underline{X}_t$ are considered; they are defined as *prediction*, *filtering* and *smoothing*. The definitions are as follows:

- Estimating $\underline{X}_t$ in terms of $\underline{Y}_0, \ldots, \underline{Y}_{t-1}$ defines the *prediction problem*;
- Estimating $\underline{X}_t$ in terms of $\underline{Y}_0, \ldots, \underline{Y}_t$ defines the *filtering problem*;
- Estimating $\underline{X}_t$ in terms of $\underline{Y}_0, \ldots, \underline{Y}_n, \ n > t,$ defines the *smoothing problem*.

Kalman filtering deals with (recursive) best linear estimation of $\underline{X}_t$ in terms of observations of $\underline{Y}_1, \underline{Y}_2, \ldots$ and a random vector $\underline{Y}_0$ which is uncorrelated with $\underline{V}_t$ and $\underline{W}_t$ for all $t \geq 1$. *Kalman recursions* is another term for *Kalman filter*.

Recall the state-space model defined by Equation (10.1).

Notation For a random vector $\underline{V}$, the notation $P_t(\underline{V}) = P_{\mathcal{F}_{0:t}^{(Y)}}(\underline{V})$ will be used.
The following theorem gives the solution to the one-step prediction problem.

Theorem 11.1 (Kalman Prediction)**.** *Consider the system defined by Equations* (10.1) *and* (10.2)*. The one-step predictors* $\widehat{\underline{X}}_t := P_{t-1}(\underline{X}_t)$ *and their error covariance matrices*

$$\Omega_t := \mathbb{E}[(\underline{X}_t - \widehat{\underline{X}}_t)(\underline{X}_t - \widehat{\underline{X}}_t)^t]$$

are uniquely determined by the initial conditions

$$\widehat{\underline{X}}_1 = P_0(\underline{X}_1), \quad \Omega_1 := \mathbb{E}[(\underline{X}_1 - \widehat{\underline{X}}_1)(\underline{X}_1 - \widehat{\underline{X}}_1)^t]$$

and the recursions, for $t = 1, \ldots,$

$$\widehat{\underline{X}}_{t+1} = G_t\widehat{\underline{X}}_t + \Theta_t\Delta_t^{-1}(\underline{Y}_t - F_t\widehat{\underline{X}}_t) \tag{11.1}$$

where

$$\widehat{X}_1 = \mathbb{P}(X_1|Y_0), \qquad \Omega_1 = \mathbb{E}[(X_1 - \widehat{X}_1)(X_1 - \widehat{X}_1)']$$

and:

$$\begin{aligned}
\Delta_t &= F_t \Omega_t F_t' + R_t \\
\Theta_t &= G_t \Omega_t F_t' + S_t \\
\Omega_{t+1} &= G_t \Omega_t G_t' + Q_t - \Theta_t \Delta_t^{-1} \Theta_t' \qquad (11.2)
\end{aligned}$$

and Δ_t^{-1} *is a generalised inverse of* Δ_t.

Definition 11.2 (The Kalman Gain)**.** *The matrix* $\Theta_t \Delta_t^{-1}$ *is called the* Kalman gain.

Example 11.1 (Local Trend Model)**.**

We now show the recursive steps for the local trend model,

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

Here $F_t = 1$, $G_t = 1$, $S_t = 0$ for each t and hence $\Omega_t = \Theta_t$. The Kalman recursion algorithm in this setting is:

$$\begin{cases} \widehat{\mu}_{t+1} - \widehat{\mu}_t = K_t(Y_t - \widehat{\mu}_t) \\ \Delta_t = \Omega_t + \sigma_e^2 \\ \Omega_{t+1} = \Omega_t + \sigma_\eta^2 - \frac{\Omega_t^2}{\Delta_t} \end{cases}$$

Here

$$\Omega_{t+1} - \Omega_t = \sigma_\eta^2 - \frac{\Omega_t^2}{\Delta_t}$$

so that:

$$\begin{cases} \Omega_{t+1} = \Omega_t(1 - K_t) + \sigma_\eta^2 \\ \Delta_t = \Omega_t + \sigma_e^2 \end{cases}$$

where the Kalman gain is $K_t := \frac{\Omega_t}{\Delta_t}$ and $\widehat{\mu}_t = \mathbb{E}[\mu_t|\mathcal{F}_{t-1}]$.

One advantage of state-space models is that missing values can be handled with relative ease. Suppose that the observations $(Y_t)_{t=l+1}^{t=l+h}$ are missing, where $h \geq 1$ and $1 \leq l \leq T$. There are several ways to handle missing variables. Here we discuss a method that keeps the original time scale and model form.

From the equation, it follows that for $t \geq l$,

$$\mu_t = \mu_l + \sum_{j=l+1}^{t} \eta_{j-1}.$$

For $s < t$, use the notation $\widehat{\mu}_{t|s} = \mathbb{E}[\mu_t|\mathcal{F}_s]$. Also, from the definition:

$$\Omega_t = \mathbb{E}[(\mu_t - \widehat{\mu}_t)^2] = \mathbb{E}[\mathrm{Var}(\mu_t|\mathcal{F}_{t-1})].$$

For $t \in \{l+1, \ldots, l+h\}$,

$$\begin{cases} \mathbb{E}[\mu_t|\mathcal{F}_{t-1}] = \mathbb{E}[\mu_t|\mathcal{F}_l] = \widehat{\mu}_{l+1|l} = \widehat{\mu}_{l+1} \\ \mathrm{Var}(\mu_t|\mathcal{F}_{t-1}) = \mathrm{Var}(\mu_t|\mathcal{F}_l) = \mathrm{Var}(\mu_{l+1}|\mathcal{F}_l) + (t-l-1)\sigma_\eta^2. \end{cases}$$

Hence

$$\mu_{t|t-1} = \mu_{t-1|t-2} \qquad \Omega_{t|t-1} = \Omega_{t-1|t-2} + \sigma_\eta^2 \qquad t = l+2, \ldots, l+h.$$

The Kalman recursion can still be applied by taking $K_t = 0$ for $t = l+1, \ldots, l+h$. This is rather natural; when Y_t is missing, there is no new innovation or new Kalman gain.

Proof of Theorem 11.1 We prove this in the case where $S_t = 0$; the innovations for 'observed' and 'hidden' processes are independent. Let the *innovations*, $\underline{I}_t$, be defined by $\underline{I}_0 := \underline{Y}_0$ and

$$\underline{I}_t := \underline{Y}_t - P_{t-1}(\underline{Y}_t) = \underline{Y}_t - F_t\widehat{\underline{X}}_t = F_t(\underline{X}_t - \widehat{\underline{X}}_t) + \underline{W}_t \qquad t \in \mathbb{N}$$

It follows from (**??**) that $\{\underline{I}_t : t \in \mathbb{N}\}$ are orthogonal. Furthermore, $\underline{Y}_0, \ldots, \underline{Y}_t$ and $\underline{Y}_0, \ldots, \underline{Y}_{t-1}, \underline{I}_t$ contain the same information, or span the same Hilbert space. It follows that

$$P_t(\underline{X}) = P(\underline{X} \mid \underline{Y}_0, \ldots, \underline{Y}_{t-1}, \underline{I}_t) = P_{t-1}(\underline{X}) + P(\underline{X} \mid \underline{I}_t), \tag{11.3}$$

where the last equality follows from Equation (**??**).

Note The following notation is used: $P(X|Z)$ denotes the projection of X onto the space spanned by Z (where Z is a collection of random variables). When $Y = \{Y_0, Y_1, \ldots, Y_t\}$, this is abbreviated to: $P(X|Y_0, Y_1, \ldots, Y_t) =: P_t(X)$.

Consequently:

$$\begin{aligned} \widehat{\underline{X}}_{t+1} &= P_t(\underline{X}_{t+1}) = P_{t-1}(\underline{X}_{t+1}) + P(\underline{X}_{t+1} \mid \underline{I}_t) \\ &= P_{t-1}(G_t\underline{X}_t + \underline{V}_{t+1}) + \mathbb{E}\left[\underline{X}_{t+1}\underline{I}_t^t\right]\left(\mathbb{E}\left[\underline{I}_t\underline{I}_t^t\right]\right)^{-1}\underline{I}_t \\ &= P_{t-1}(G_t\underline{X}_t) + \mathbb{E}\left[\underline{X}_{t+1}\underline{I}_t^t\right]\left(\mathbb{E}\left[\underline{I}_t\underline{I}_t^t\right]\right)^{-1}\underline{I}_t. \end{aligned}$$

Here the basic idea of projection in a Hilbert space is used. Let H be a Hilbert space, with $x \in H$ and $y \in H$, then

$$P(x|y) = \langle x, \frac{y}{\|y\|}\rangle\frac{y}{\|y\|} = \frac{\langle x, y\rangle}{\|y\|^2}y$$

where $\langle .,. \rangle$ denotes the inner product and $\|.\|$ denotes the norm for the Hilbert space. Recall $\Delta_t := F_t \Omega_t F_t' + R_t$ so that $\Delta_t = \mathbb{E}\left[\underline{I}_t \underline{I}_t^t\right]$. From the above, it follows that:

$$\Delta_t := F_t \Omega_t F_t' + R_t = F_t \mathbb{E}\left[\left(\underline{X}_t - \widehat{\underline{X}}_t\right)\left(\underline{X}_t - \widehat{\underline{X}}_t\right)'\right] F_t' + \mathbb{E}\left[\underline{W}_t \underline{W}_t'\right] = \mathbb{E}\left[\underline{I}_t \underline{I}_t'\right]$$

and (when $S_t = 0$), $\Theta_t = G_t \Omega_t F_t'$ so that:

$$\begin{aligned} \Theta_t &:= G_t \Omega_t F_t' \\ &= \mathbb{E}\left[(G_t \underline{X}_t)(\underline{X}_t - \widehat{\underline{X}}_t)' F_t'\right] \\ &= \mathbb{E}[(G_t \underline{X}_t + \underline{V}_{t+1})((\underline{X}_t - \widehat{\underline{X}}_t)' F_t' + \underline{W}_t')] \\ &= = \mathbb{E}\left[\underline{X}_{t+1} \underline{I}_t'\right] \end{aligned}$$

To establish Equation (11.2), note that

$$\Omega_{t+1} = \mathbb{E}\left[\underline{X}_{t+1} \underline{X}_{t+1}'\right] - \mathbb{E}\left[\widehat{\underline{X}}_{t+1} \widehat{\underline{X}}_{t+1}'\right].$$

From this,

$$\mathbb{E}\left[\underline{X}_{t+1} \underline{X}_{t+1}'\right] = \mathbb{E}[(G_t \underline{X}_t + \underline{V}_{t+1})(G_t \underline{X}_t + \underline{V}_{t+1})'] = G_t \mathbb{E}\left[\underline{X}_t \underline{X}_t'\right] G_t' + Q_{t+1}$$

and

$$\begin{aligned} \mathbb{E}\left[\widehat{\underline{X}}_{t+1} \widehat{\underline{X}}_{t+1}'\right] &= \mathbb{E}\left[(G_t \widehat{\underline{X}}_t + \Theta_t \Delta_t^{-1} \underline{I}_t)(G_t \widehat{\underline{X}}_t + \Theta_t \Delta_t^{-1} \underline{I}_t)'\right] \\ &= G_t \mathbb{E}\left[\widehat{\underline{X}}_t \widehat{\underline{X}}_t'\right] G_t' + \Theta_t \Delta_t^{-1} \Delta_t \Delta_t^{-1} \Theta_t = G_t \mathbb{E}\left[\widehat{\underline{X}}_t \widehat{\underline{X}}_t'\right] G_t' + \Theta_t \Delta_t^{-1} \Theta_t. \end{aligned}$$

From this, it follows that:

$$\Omega_{t+1} = G_t \left(\mathbb{E}\left[\underline{X}_t \underline{X}_t'\right] - \mathbb{E}\left[\widehat{\underline{X}}_t \widehat{\underline{X}}_t'\right]\right) G_t' + Q_t - \Theta_t \Delta_t^{-1} \Theta_t = G_t \Omega_t G_t' + Q_{t+1} - \Theta_t \Delta_t^{-1} \Theta_t,$$

as required. □

Theorem 11.3 (Kalman Filtering). *The filtered estimates $\underline{X}_{t|t} := P_t(\underline{X}_t)$ and the error covariance matrices*

$$\Omega_{t|t} := \mathbb{E}\left[(\underline{X}_t - \underline{X}_{t|t})(\underline{X}_t - \underline{X}_{t|t})^t\right]$$

are determined by the relations

$$X_{t|t} = P_{t-1}(\underline{X}_t) + \Omega_t F_t^t \Delta_t^{-1}(\underline{Y}_t - F_t \widehat{\underline{X}}_t)$$

and

$$\Omega_{t|t+1} = \Omega_t - \Omega_t F_t^t \Delta_t^{-1} F_t \Omega_t^t.$$

Proof From Equation (11.3), it follows that

$$P_t\underline{X}_t = P_{t-1}\underline{X}_t + M\underline{I}_t,$$

where

$$M = \mathbb{E}\left[\underline{X}_t\underline{I}_t^t\right]\left(\mathbb{E}\left[\underline{I}_t\underline{I}_t^t\right]\right)^{-1} = \mathbb{E}\left[\underline{X}_t(F_t(\underline{X}_t - \widehat{\underline{X}}_t) + W_t)^t\right]\Delta_t^{-1} = \Omega_t F_t \Delta_t^{-1}.$$

It follows that:

$$\underline{X}_t - P_{t-1}\underline{X}_t = \underline{X}_t - P_t\underline{X}_t + P_t\underline{X}_t - P_{t-1}\underline{X}_t = \underline{X}_t - P_t\underline{X}_t + M\underline{I}_t.$$

Now use $\underline{X}_t - P_t\underline{X}_t \perp M\underline{I}_t$ to obtain:

$$\Omega_t = \Omega_{t|t} + \Omega_t F_t \Delta_t^{-1} F_t' \Omega_t^t$$

as required. □

Theorem 11.4 (Kalman Fixed Point Smoothing)**.** *The smoothed estimates* $\underline{X}_{t|n} := P_n(\underline{X}_t)$ *and the error covariance matrices*

$$\Omega_{t|n} := \mathbb{E}[(\underline{X}_t - \underline{X}_{t|n})(\underline{X}_t - \underline{X}_{t|n})']$$

are determined for fixed t *by the recursions, which can be solved successively for* $n = t, t+1, \ldots$:

$$\begin{aligned} P_n(\underline{X}_t) &= P_{n-1}(\underline{X}_t) + \Omega_{t,n}F_n^t\Delta_n^{-1}(\underline{Y}_n - F_n\widehat{\underline{X}}_n), \\ \Omega_{t,n+1} &= \Omega_{t,n}\left(G_n - \Theta_n\Delta_n^{-1}F_n\right)^t, \\ \Omega_{t|n} &= \Omega_{t|n-1} - \Omega_{t,n}F_n^t\Delta_n^{-1}F_n\Omega_{t.n}^t, \end{aligned}$$

with initial conditions $P_{t-1}(\underline{X}_t) = \widehat{\underline{X}}_t$ *and* $\Omega_{t.t} = \Omega_{t|t-1} = \Omega_t$ *found from Kalman prediction.*

Proof Again, from Equation (11.3), $P_n\underline{X}_t = P_{n-1}\underline{X}_t + C\underline{I}_n$ where

$$\underline{I}_n = F_n(\underline{X}_n - \widehat{\underline{X}}_n) + \underline{W}_n.$$

Using the fact that

$$P(\underline{X}|\underline{Y}) = M\underline{Y}$$

where

$$M = \mathbb{E}[\underline{XY}^t]\left(\mathbb{E}\left[\underline{YY}^t\right]\right)^{-1},$$

$\left(\mathbb{E}\left[\underline{YY}^t\right]\right)^{-1}$ any generalised inverse of $\mathbb{E}\left[\underline{YY}^t\right]$, it follows that

$$C = \mathbb{E}\left[\underline{X}_t(F_n(\underline{X}_n - \widehat{X}_n) + \underline{W}_n)^t\right]\left(\mathbb{E}\left[\underline{I}_n\underline{I}_n^t\right]\right)^{-1} = \Omega_{t,n}F_n^t\Delta_n^{-1}. \quad (11.4)$$

It now follows using

$$\begin{cases} \underline{Y}_t = F_t \underline{X}_t + \underline{W}_t \\ \underline{X}_{t+1} = G_t \underline{X}_t + \underline{V}_{t+1} \end{cases}$$

and Equation (11.1), which may be re-written as:

$$P_t \underline{X}_{t+1} = G_t P_{t-1} \underline{X}_t + \Theta_t \Delta_t^{-1} (\underline{Y}_t - P_{t-1} \underline{Y}_t)$$

together with the orthogonality of $\underline{V}_{n+1}$ and $\underline{W}_n$ with $\underline{X}_t - \widehat{\underline{X}}_t$ and the definition of $\Omega_{t,n}$ that

$$\Omega_{t,n+1} = \mathbb{E}\left[(\underline{X}_t - \widehat{\underline{X}}_t)(\underline{X}_n - \widehat{\underline{X}}_n)^t\right](G_n - \Theta_n \Delta_n^{-1} F_n)^t = \Omega_{t,n}(G_n - \Theta_n \Delta_n^{-1} F_n)^t$$

thus establishing the equation for $\Omega_{t,n+1}$. To establish the equation for $\Omega_{t|n}$,

$$\underline{X}_t - P_n \underline{X}_t = \underline{X}_t - P_{n-1} \underline{X}_t - C \underline{I}_n$$

Using Equation (11.4) and the fact that $\underline{X}_t - P_n \underline{X}_t \perp \underline{I}_n$, it follows that

$$\Omega_{t|n} = \Omega_{t|n-1} - \Omega_{t,n} F_n^t \Delta_n^{-1} \Omega_{t,n}^t \qquad n = t, t+1, \ldots$$

as required. □