Financial Econometrics

Linear State Space Models

In Choi

Sogang University

November 2, 2022

- Model

Observation equation:

$$
\begin{aligned}
y_t &= \underset{p\times m}{Z_t}\,\alpha_t + \varepsilon_t,\ \ \varepsilon_t \sim N(0, H_t), (t = 1, ..., n),\\
\textbf{Transition equation} &:\\
\alpha_{t+1} &= T_t\alpha_t + R_t\eta_t,\ \ \eta_t \sim N(0, Q_t),\\
\alpha_1 &\sim N(a_1, P_1),
\end{aligned}
$$

where y_t is a $p \times 1$ vector of observations called the **observation vector** and α_t is an unobserved $m \times 1$ vector called the **state vector**.

- Assumptions
  - The matrices Z_t, T_t, R_t, H_t and Q_t are known; a_1 and P_1 are also known.
  - The error terms ε_t and η_t are serially independent and independent of each other at all time points.
  - α_1 is independent of $\{\varepsilon_t\}$ and $\{\eta_t\}$.

- Model

$$\begin{aligned} y_t &= \alpha_t + \varepsilon_t, \ \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \\ \alpha_{t+1} &= \alpha_t + \eta_t, \ \eta_t \sim N(0, \sigma_\eta^2), \\ \alpha_1 &\sim N(a_1, P_1), \end{aligned}$$

  where ε_t's and η_t's are all mutually independent and are independent of α_1.
- If $\sigma_\eta^2 > 0$, y_t is the sum of a **random walk** and a noise term.
- If $\sigma_\eta^2 = 0$, $\alpha_{t+1} = \alpha_t = ... = \alpha_1$. Thus, y_t is the sum of a **constant** and a noise term.

- Model

$$
\begin{aligned}
y_t &= \mu_t + \varepsilon_t,\ \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \\
\mu_{t+1} &= \mu_t + \nu_t + \eta_t,\ \eta_t \sim N(0, \sigma_\eta^2), \\
\nu_{t+1} &= \nu_t + \varsigma_t,\ \varsigma_t \sim N(0, \sigma_\varsigma^2),\ (\varsigma:\ \text{sigma}) \\
\alpha_1 &\sim N(a_1, P_1).
\end{aligned}
$$

- This can be written in state space form as

$$
\begin{aligned}
y_t &= (1\ 0)\begin{pmatrix} \mu_t \\ \nu_t \end{pmatrix} + \varepsilon_t, \\
\begin{pmatrix} \mu_{t+1} \\ \nu_{t+1} \end{pmatrix} &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{pmatrix} \mu_t \\ \nu_t \end{pmatrix} + \begin{pmatrix} \eta_t \\ \varsigma_t \end{pmatrix}.
\end{aligned}
$$

- If $\sigma_\eta^2 = \sigma_\varsigma^2 = 0$, $\nu_{t+1} = \nu_t$ and $\mu_{t+1} = \mu_t + \nu_t$. Thus, $\nu_{t+1} = \nu_t = ... = \nu_1$ and $\mu_{t+1} = \mu_t + \nu_1 = \mu_1 + t\nu_1$. So the model reduces to the **deterministic linear trend** plus noise model.
- If $\sigma_\eta^2 = 0$ and $\sigma_\varsigma^2 > 0$, ν_{t+1} is a random walk and μ_{t+1} is the sum of the ramdom walk. Thus, the model becomes the **integrated random walk**.
- If $\sigma_\eta^2 > 0$ and $\sigma_\varsigma^2 = 0$, $\mu_{t+1} = \mu_t + \nu_1 + \eta_t$. So the model becomes the **deterministic linear trend** plus **random walk** model.

- Model (local linear trend + seasonality)

$$
\begin{aligned}
y_t &= \mu_t + \gamma_t + \varepsilon_t, \;\; \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \\
\mu_{t+1} &= \mu_t + \nu_t + \eta_t, \;\; \eta_t \sim N(0, \sigma_\eta^2), \\
\nu_{t+1} &= \nu_t + \varsigma_t, \;\; \varsigma_t \sim N(0, \sigma_\varsigma^2)
\end{aligned}
$$

- Models for seasonality (s :# of seasons)

$$
\begin{aligned}
(i)\ \gamma_{t+1} &= -\sum_{j=1}^{s-1} \gamma_{t+1-j}; \\
(ii)\ \gamma_{t+1} &= -\sum_{j=1}^{s-1} \gamma_{t+1-j} + \omega_t,\ \ \omega_t \sim N(0, \sigma_\omega^2); \\
(iii)\ \gamma_{j,t+1} &= \gamma_{j,t} + \omega_{jt},\ \ t = (i-1)s + j, (i = 1, 2, ...;\ j = 1, ..., s) \\
\text{with} \sum_{j=1}^{s} \gamma_{j,t} &= 0 \text{ for any } t. \text{ (quasi-random walk)}
\end{aligned}
$$

- For (ii), take the state vector as

$$\alpha_t = (\mu_t, \nu_t, \gamma_t, \ldots, \gamma_{t-s+2})'$$

and define the system matrices accordingly.

- ARMA(2,1) model
  Transition equation

$$\begin{bmatrix} y_{t+1} \\ \phi_2 y_t + \theta_1 \zeta_{t+1} \end{bmatrix} = \begin{bmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{bmatrix} \begin{bmatrix} y_t \\ \phi_2 y_{t-1} + \theta_1 \zeta_t \end{bmatrix} + \begin{pmatrix} 1 \\ \theta_1 \end{pmatrix} \zeta_{t+1}$$

$$(\zeta \quad : \quad \text{zeta})$$

Observational equation

$$y_t = (1\ 0)\alpha_t$$

- ARIMA(2,1,1) model

$$
\begin{aligned}
\alpha_t &= \begin{bmatrix} y_{t-1} \\ y_t^* \\ \phi_2 y_{t-1}^* + \theta_1 \zeta_t \end{bmatrix}, \; y_t^* = \Delta y_t \\
y_t &= (1\ 1\ 0)\alpha_t : \text{identity relation} \\
\alpha_{t+1} &= \begin{bmatrix} 1 & 1 & 0 \\ 0 & \phi_1 & 1 \\ 0 & \phi_2 & 0 \end{bmatrix} \alpha_t + \begin{pmatrix} 0 \\ 1 \\ \theta_1 \end{pmatrix} \zeta_{t+1}
\end{aligned}
$$

- ARIMA(2,1,1) model
  The third equation means

$$
\begin{aligned}
y_t &= y_{t-1} + \Delta y_t = y_t \quad \text{: identity relation} \\
\Delta y_{t+1} &= \phi_1 \Delta y_t + \phi_2 \Delta y_{t-1} + \theta_1 \zeta_t + \zeta_{t+1} \\
\phi_2 \Delta y_t + \theta_1 \zeta_{t+1} &= \phi_2 \Delta y_t + \theta_1 \zeta_{t+1} : \ \text{identity relation}
\end{aligned}
$$

- Model

$$
\begin{aligned}
y_t &= s_t + x_t \\
s_t &= \beta s_{t-\tau} + e_t, \;\; e_t \sim iid\ (0, \sigma_e^2) \\
x_t &= \sum_{k=1}^{p} \phi_k x_{t-k} + u_t + \sum_{l=1}^{q} \theta_l u_{t-l}, \;\; u_t \sim iid\ (0, \sigma_u^2)
\end{aligned}
\tag{1}
$$

- The transition equation for the seasonal component is written as

$$\xi_{t+1} = V\xi_t + Ee_{t+1}, \tag{2}$$

where

$$\xi_t = \begin{pmatrix} s_t \\ \beta \begin{bmatrix} s_{t-\tau+1} \\ \vdots \\ s_{t-1} \end{bmatrix} \end{pmatrix}, \quad V_i = \begin{bmatrix} \mathbf{0}_{\tau-1} & I_{\tau-1} \\ \beta & \mathbf{0}'_{\tau-1} \end{bmatrix}, \quad E = \begin{pmatrix} 1 \\ \mathbf{0}_{\tau-1} \end{pmatrix},$$

- The transition equation for the random component is

$$\varsigma_{t+1} = W\varsigma_t + Uu_{t+1}, \tag{3}$$

where

$$\varsigma_t = \begin{bmatrix} x_t \\ \phi_2 x_{t-1} + \ldots + \phi_r x_{t-r+1} + \theta_1 u_t + \ldots + \theta_{r-1} u_{t-r+2} \\ \phi_3 x_{t-1} + \ldots + \phi_r x_{t-r+2} + \theta_2 u_t + \ldots + \theta_{r-1} u_{t-r+2} \\ \vdots \\ \phi_r x_{t-1} + \theta_{r-1} u_t \end{bmatrix},$$

$$W = \begin{bmatrix} \phi_1 & \\ \vdots & I_{r-1} \\ \phi_{r-1} & \\ \phi_r & \mathbf{0}'_{r-1} \end{bmatrix}, \; U_i = \begin{pmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{r-1} \end{pmatrix},$$

and $r = \max(p, q+1)$. In the special case $r = 1$, $W = \phi_1$ and $U = 1$.

- Let

$$
\begin{aligned}
\alpha_t &= \begin{bmatrix} \xi_t' & \varsigma_t' \end{bmatrix}', \\
T &= \begin{bmatrix} V & \mathbf{0}_{\tau\times r} \\ \mathbf{0}_{r\times\tau} & W \end{bmatrix}, \\
\eta_t &= \begin{pmatrix} e_{t+1} \\ u_{t+1} \end{pmatrix}, \ R = \begin{bmatrix} 1 & 0 \\ \mathbf{0}_{\tau-1} & \mathbf{0}_{\tau-1} \\ \mathbf{0}_r & U \end{bmatrix}.
\end{aligned}
$$

- Putting (2) and (3) together, we may write model (1) in state space form as

$$
\begin{aligned}
y_t &= Z\alpha_t, \;\; Z = [1 \; \mathbf{0}'_{\tau-1} \; 1 \; \mathbf{0}'_{r-1}], \\
\alpha_{t+1} &= T\alpha_t + R\eta_t, \;\; \eta_t \sim iid\;(\mathbf{0}_2, Q), \;\; Q = \begin{bmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_u^2 \end{bmatrix}.
\end{aligned}
$$

Advantages of the state space approach

- The different components that make up the series, (e.g., trend, seasonal, cycle and calendar variations, explanatory variables and interventions) are modelled **separately** before being put together in the state space model. The investigator can identify each component **separately** using the state space approach.

Derivation of the Kalman filter

- Assume that a_1 and P_1 are known. Let $Y_t = (y_1, ..., y_{t-1})'$. Our objective is to obtain

$$\begin{aligned} a_{t|t} &= E(\alpha_t \mid Y_t), P_{t|t} = Var(\alpha_t \mid Y_t) \\ a_{t+1} &= E(\alpha_{t+1} \mid Y_t), P_{t+1} = Var(\alpha_{t+1} \mid Y_t). \end{aligned}$$

- Assume

$$\alpha_t \mid Y_t \sim N(a_{t|t}, P_{t|t})$$

and

$$\alpha_{t+1} \mid Y_t \sim N(a_{t+1}, P_{t+1}).$$

- Starting with $N(a_t, P_t)$, we calculate $a_{t|t}, a_{t+1}, P_{t|t}$ and P_{t+1} from a_t and P_t recursively.

- Let
$$v_t = y_t - E(y_t \mid Y_{t-1}) = y_t - Z_t a_t \tag{4}$$
(**one-step ahead forecast error** of y_t given Y_{t-1}).
- Since $E(v_t \mid Y_{t-1}) = E(y_t - Z_t a_t \mid Y_{t-1}) = E(Z_t \alpha_t + \varepsilon_t - Z_t a_t \mid Y_{t-1}) = 0$. We have for $j = 1, ..., t-1$

$$\begin{aligned} E(v_t) &= 0 \\ Cov(y_j, v_t) &= EE\left(y_j v_t \mid Y_{t-1}\right) \\ &= E\left\{y_j E\left(v_t \mid Y_{t-1}\right)\right\} \\ &= 0. \end{aligned}$$

- When Y_{t-1} and v_t are fixed, then Y_t is fixed and vice versa.[1] Thus,

$$\begin{aligned} a_{t|t} &= E(\alpha_t \mid Y_t) = E(\alpha_t \mid Y_{t-1}, v_t), \\ a_{t+1} &= E(\alpha_{t+1} \mid Y_t) = E(\alpha_{t+1} \mid Y_{t-1}, v_t) \end{aligned}$$

[1] Recall that $y_t = v_t + E(y_t \mid Y_{t-1}) = v_t + Z_t a_t$.

Derivation of the Kalman filter

Step 1

Lemma 1 Suppose that

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma'_{xy} & \Sigma_{yy} \end{pmatrix} \right).$$

Then,

$$x \mid y \sim N\left(\mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y), \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma'_{xy} \right).$$

- Apply Lemma 1 to the conditional joint distribution of α_t and v_t given Y_{t-1}. Taking x and y in Lemma 1 as α_t and v_t, we obtain

$$\begin{aligned} a_{t|t} &= E(\alpha_t \mid Y_t) = E(\alpha_t \mid Y_{t-1}, v_t) \\ &= E(\alpha_t \mid Y_{t-1}) + Cov(\alpha_t, v_t \mid Y_{t-1}) Var(v_t \mid Y_{t-1})^{-1} v_t. \end{aligned}$$

But

$$\begin{aligned} Cov(\alpha_t, v_t \quad | \quad Y_{t-1}) &= E\left(\alpha_t(Z_t\alpha_t + \varepsilon_t - Z_t a_t)' \mid Y_{t-1}\right) \\ &= E\left(\alpha_t(\alpha_t - a_t)'Z_t' \mid Y_{t-1}\right) \\ &= P_t Z_t' \ (\text{Recall } P_t = Var(\alpha_t \mid Y_{t-1})) \end{aligned}$$

and

$$\begin{aligned} Var(v_t \quad | \quad Y_{t-1}) &= Var(Z_t\alpha_t + \varepsilon_t - Z_t a_t \mid Y_{t-1}) \\ &= Z_t P_t Z_t' + H_t \\ &= F_t, \ \text{say.} \end{aligned} \tag{5}$$

Thus

$$a_{t|t} = a_t + P_t Z_t' F_t^{-1} v_t. \tag{6}$$

- Using Lemma 1, we obtain

$$
\begin{aligned}
P_{t|t} &= Var(\alpha_t \mid Y_t) = Var(\alpha_t \mid Y_{t-1}, v_t) \qquad (7)\\
&= Var(\alpha_t \mid Y_{t-1}) \\
-Cov(\alpha_t, v_t &\mid Y_{t-1}) Var(v_t \mid Y_{t-1})^{-1} Cov(\alpha_t, v_t \mid Y_{t-1})' \\
&= P_t - P_t Z_t' F_t^{-1} Z_t P_t'.
\end{aligned}
$$

- Relations (6) and (7) are called the **updating step** of the Kalman filter.

Derivation of the Kalman filter

Step 2

- Now develop recursion for α_{t+1} and P_{t+1}.

$$a_{t+1} = E(T_t\alpha_t + R_t\eta_t \mid Y_t) = T_t a_{t|t} \tag{8}$$

$$\begin{aligned} P_{t+1} &= Var(T_t\alpha_t + R_t\eta_t \mid Y_t) \\ &= T_t Var(\alpha_t \mid Y_t) T_t' + R_t Q_t R_t' \\ &= T_t P_{t|t} T_t' + R_t Q_t R_t' \end{aligned} \tag{9}$$

- Substituting (6) into (8) gives

$$\begin{aligned} a_{t+1} &= T_t(a_t + P_t Z_t' F_t^{-1} v_t) \\ &= T_t a_t + K_t v_t, \end{aligned} \tag{10}$$

where $K_t = T_t P_t Z_t' F_t^{-1}$ (called the **Kalman gain**).

- Substituting (7) into (9) gives

$$\begin{aligned} P_{t+1} &= T_t\left(P_t - P_t Z_t' F_t^{-1} Z_t P_t'\right) T_t' + R_t Q_t R_t' \\ &= T_t P_t (T_t - K_t Z_t)' + R_t Q_t R_t' \end{aligned} \tag{11}$$

- Summary

$$
\begin{aligned}
v_t &= y_t - Z_t a_t, \;\; F_t = Z_t P_t Z_t' + H_t, \\
a_{t|t} &= a_t + P_t Z_t' F_t^{-1} v_t, \;\; P_{t|t} = P_t - P_t Z_t' F_t^{-1} Z_t P_t', \\
a_{t+1} &= T_t a_t + K_t v_t, \;\; P_{t+1} = T_t P_t (T_t - K_t Z_t)' + R_t Q_t R_t'.
\end{aligned}
$$

- Although the results are obtained under the assumption of normality, they have a wider validity in the sense of minimum variance linear unbiased estimation when the variables involved are not normally distributed. (Use Lemma 2 in Section 3 of DK.)

- Define the **state estimation error** as

$$x_t = \alpha_t - a_t \text{ with } Var(x_t) = P_t.$$

- The one-step ahead forecast error v_t (called also **innovation**) can be written as

$$\begin{aligned} v_t &= y_t - E(y_t \mid Y_{t-1}) = y_t - Z_t a_t \\ &= Z_t \alpha_t + \varepsilon_t - Z_t a_t \\ &= Z_t x_t + \varepsilon_t. \end{aligned} \tag{12}$$

- Thus, the recursive relation for state esimation error is given as

$$
\begin{aligned}
x_{t+1} &= \alpha_{t+1} - a_{t+1} \\
&= T_t \alpha_t + R_t \eta_t - T_t a_t - K_t v_t \\
&= T_t x_t + R_t \eta_t - K_t Z_t x_t - K_t \varepsilon_t \\
&= L_t x_t + R_t \eta_t - K_t \varepsilon_t, \ (L_t = T_t - K_t Z_t),
\end{aligned} \tag{13}
$$

where the second equality employs relation (10).

State smoothing

- The objective of **state smoothing**
  Derive the conditional density of α_t given the entire series $y_1, \ldots, y_n$.

- The operation of calculating $\hat{\alpha}_t = E(\alpha_t \mid Y_n)$ is called **state smoothing**.
- Let $v_{t:n} = (v_t', ..., v_n')'$. Y_n is fixed when Y_{t-1} and $v_{t:n}$ are fixed. Calculate the conditional mean of α_t given Y_{t-1} and $v_{t:n}$. Using Lemma 1, we obtain

$$\begin{aligned} \hat{\alpha}_t &= E(\alpha_t \mid Y_n) = E(\alpha_t \mid Y_{t-1}, v_{t:n}) \\ &= a_t + \sum_{j=t}^{n} Cov(\alpha_t, v_j \mid Y_{t-1}) F_j^{-1} v_j, \end{aligned} \tag{14}$$

where $F_j = Var(v_j \mid Y_{t-1})$.

- Relations (12) and (13) provide

$$
\begin{aligned}
Cov(\alpha_t, v_j \quad | \quad Y_{t-1}) &= E(\alpha_t v_j' \mid Y_{t-1}) \\
&= E\left[\alpha_t (Z_j x_j + \varepsilon_j)' \mid Y_{t-1}\right] \\
&= E\left(\alpha_t x_j' \mid Y_{t-1}\right) Z_j', \; j = t, \ldots, n. \qquad (15)
\end{aligned}
$$

- Moreover, (recall $x_{t+1} = L_t x_t + R_t \eta_t - K_t \varepsilon_t$)

$$
\begin{aligned}
E\left(\alpha_t x_t' \mid Y_{t-1}\right) &= E\left(\alpha_t(\alpha_t - a_t)' \mid Y_{t-1}\right) = P_t, \\
E\left(\alpha_t x_{t+1}' \mid Y_{t-1}\right) &= E\left[\alpha_t(L_t x_t + R_t \eta_t - K_t \varepsilon_t)' \mid Y_{t-1}\right] = P_t L_t', \\
E\left(\alpha_t x_{t+2}' \mid Y_{t-1}\right) &= P_t L_t' L_{t+1}' \\
&\vdots \\
E\left(\alpha_t x_n' \mid Y_{t-1}\right) &= P_t L_t' \ldots L_{n-1}'. \qquad (16)
\end{aligned}
$$

- When $t = n$, $L'_t \ldots L'_{n-1} = I_m$.[2] When $t = n-1$, $L'_t \ldots L'_{n-1} = L'_{n-1}$.[3]

---

[2] $E(\alpha_n x'_n \mid Y_{n-1}) = P_n$

[3]

$$\begin{aligned} E(\alpha_{n-1} x'_n \quad &| \quad Y_{n-2}) \\ &= E(\alpha_{n-1}(\alpha_n - a_n)' \mid Y_{n-2}) \\ &= E(\alpha_{n-1}(L_{n-1}x_{n-1} + R_{n-1}\eta_{n-1} - K_{n-1}\varepsilon_{n-1})' \mid Y_{n-2}) \\ &= P_{n-1}L'_{n-1} \end{aligned}$$

- Substituting (15) and (16) into (14), we have

$$\hat{\alpha}_t = a_t + P_t r_{t-1}, \tag{17}$$

where $r_{n-1} = Z_n' F_n^{-1} v_n$ and

$$r_{t-1} = Z_t' F_t^{-1} v_t + L_t' Z_{t+1}' F_{t+1}^{-1} v_{t+1} + ... + L_t' L_{t+1}' ... L_{n-1}' Z_n' F_n^{-1} v_n \tag{18}$$

for $t = 1, ..., n-1$ and $r_n = 0$.

- $\{r_t\}$ satisfies the backward recursion

$$r_{t-1} = Z_t' F_t^{-1} v_t + L_t' r_t, \ \ t = n, ..., 1 \tag{19}$$

with $r_n = 0$.

- Applying Lemma 1 to the conditional joint distribution of α_t and $v_{t:n}$ given Y_{t-1} and using (15) and (16), we have

$$
\begin{aligned}
V_t &= Var(\alpha_t \mid Y_{t-1}, v_{t:n}) = P_t \\
-\sum_{j=t}^{n} Cov(\alpha_t, v_j &\mid Y_{t-1}) F_j^{-1} Cov(\alpha_t, v_j \mid Y_{t-1})' \\
&= P_t - P_t N_{t-1} P_t,
\end{aligned}
$$

where

$$
\begin{aligned}
N_{t-1} &= Z_t' F_t^{-1} Z_t + L_t' Z_{t+1}' F_{t+1}^{-1} Z_{t+1} L_t + \ldots \\
&\quad + L_t' L_{t+1}' \ldots L_{n-1}' Z_n' F_n^{-1} Z_n L_{n-1} \ldots L_t.
\end{aligned}
$$

- The sequence $\{N_t\}$ satisfies the recursion

$$N_{t-1} = Z_t' F_t^{-1} Z_t + L_t' N_t L_t, \;\; t = n, ..., 1$$

with $N_n = 0$.

- Summary

$$
\begin{aligned}
r_{t-1} &= Z_t' F_t^{-1} v_t + L_t' r_t, \;\; N_{t-1} = Z_t' F_t^{-1} Z_t + L_t' N_t L_t, \\
\hat{\alpha}_t &= a_t + P_t r_{t-1}, \;\; V_t = P_t - P_t N_{t-1} P_t.
\end{aligned}
$$

for $t = n, ..., 1$ with $r_n = 0$ and $N_n = 0$.

Missing observations

- Suppose that $y_\tau, ..., y_{\tau^*}$ are missing.
- For $t = \tau, ..., \tau^* - 1$, we have (note: use the fact $Y_t = Y_{\tau-1}$)

$$
\begin{aligned}
a_{t|t} &= E(\alpha_t \mid Y_t) = E(\alpha_t \mid Y_{t-1}) = a_t, \\
P_{t|t} &= Var(\alpha_t \mid Y_t) = Var(\alpha_t \mid Y_{t-1}) = P_t \\
a_{t+1} &= E(\alpha_{t+1} \mid Y_t) = E(T_t\alpha_t + R_t\eta_t \mid Y_{t-1}) = T_t a_t, \\
P_{t+1} &= Var(\alpha_{t+1} \mid Y_t) = Var(T_t\alpha_t + R_t\eta_t \mid Y_{t-1}) \\
&= T_t P_{\tau-1} T_t{}' + R_t Q_t R_t'.
\end{aligned}
$$

- That is, put $Z_t = 0$ for $t = \tau, ..., \tau^* - 1$ in applying the Kalman filter and smoother.

Forecasting

- The minimum mean square error forecast of y_{n+j} $(j = 1, \ldots, J)$ given Y_n is the conditional mean $\bar{y}_{n+j} = E(y_{n+j} \mid Y_n)$.
- For $j = 1$,

$$\bar{y}_{n+1} = Z_{n+1} E(\alpha_{n+1} \mid Y_n) = Z_{n+1} a_{n+1}$$

and

$$\bar{F}_{n+1} = Z_{n+1} P_{n+1} Z'_{n+1} + H_{n+1}.$$

Note that a_{n+1} and P_{n+1} can be calculated using the Kalman filter.

Forecasting

- For $j = 2, ..., n$,

$$\bar{y}_{n+j} = Z_{n+j} E(\alpha_{n+j} \mid Y_n) = Z_{n+j} \bar{a}_{n+j}$$

  with

$$\bar{F}_{n+1} = Z_{n+j} \bar{P}_{n+j} Z'_{n+j} + H_{n+j}.$$

- The recursive relation for $\bar{a}_{n+j}$ is

$$\bar{a}_{n+j+1} = T_{n+j} \bar{a}_{n+j}$$

  for $j = 1, ..., J-1$ with $\bar{a}_{n+1} = a_{n+1}$.

Forecasting

- Also

$$\bar{P}_{n+j+1} = T_{n+j}\bar{P}_{n+j}T'_{n+j} + R_{n+j}Q_{n+j}R'_{n+j}$$

  with $\bar{P}_{n+1} = P_{n+1}$.
- These show that the recursions for $\bar{a}_{n+j}$ and $\bar{P}_{n+j+1}$ are the same as recursion for a_{n+j} and P_{n+j} of the Kalman filter provided that we take $Z_{n+j} = 0$ for $j = 1, ..., J-1$.

- We develop methods of starting up the series when a_1 and P_1 are unknown; the process is called **initialisation**.
- A general model for the initial state vector α_1

$$
\begin{aligned}
\alpha_1 &= a + A\delta + R_0\eta_0, \;\; \eta_0 \sim N(0, Q_0) \\
a &: m \times 1 \text{ (known, usually zero vector)} \\
A &: m \times q, \;\; \delta : q \times 1 \text{ (unknown)}, \; R_0 : m \times (m-q)
\end{aligned}
$$

- The objective of this representation is to separate out α_1 into a constant part a, a nonstationary part $A\delta$ and a stationary part $R_0\eta_0$.
- The diffuse initialization (introduced later) is used for δ only.

- In most cases, the columns of A and R_0 are taken from those of I_m, and $A'R_0 = 0$. In some cases, $R_0 = 0$ and $A = I_m$.

Example

Consider

$$
\begin{aligned}
y_t &= \mu_t + \rho_t + \varepsilon_t, \;\; \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \\
\mu_{t+1} &= \mu_t + \nu_t + \xi_t, \;\; \xi_t \sim N(0, \sigma_\xi^2), \\
\nu_{t+1} &= \nu_t + \zeta_t, \;\; \zeta_t \sim N(0, \sigma_\zeta^2), \\
\rho_{t+1} &= \phi\rho_t + \tau_t, \;\; \tau_t \sim n(0, \sigma_\tau^2), \\
|\phi| &< 1.
\end{aligned}
$$

Example

(Continued) In state-space format, this is

$$
\begin{aligned}
y_t &= \begin{bmatrix} 1 & 0 & 1 \end{bmatrix} \begin{pmatrix} \mu_t \\ \nu_t \\ \rho_t \end{pmatrix} + \varepsilon_t, \\
\begin{pmatrix} \mu_{t+1} \\ \nu_{t+1} \\ \rho_{t+1} \end{pmatrix} &= \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \phi \end{bmatrix} \begin{pmatrix} \mu_t \\ \nu_t \\ \rho_t \end{pmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} \xi_t \\ \zeta_t \\ \tau_t \end{pmatrix}.
\end{aligned}
$$

Example

(Continued) Thus

$$
\begin{aligned}
a &= \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \ A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \ \delta = \begin{pmatrix} \mu_1 \\ \nu_1 \end{pmatrix}, \ R_0 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \\
\eta_0 &= \rho_1, \ Q_0 = \sigma_\tau^2/(1-\phi^2). \\
\text{Note} &: \begin{pmatrix} \mu_1 \\ \nu_1 \\ \rho_1 \end{pmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{pmatrix} \mu_1 \\ \nu_1 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \rho_1.
\end{aligned}
$$

Example

Consider

$$
\begin{aligned}
y_t &= \mu_t + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \\
\mu_{t+1} &= \mu_t + \nu_t + \xi_t, \xi_t \sim N(0, \sigma_\xi^2), \\
\nu_{t+1} &= \nu_t + \zeta_t, \ \zeta_t \sim N(0, \sigma_\zeta^2).
\end{aligned}
$$

$$
a = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \ A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \ \delta = \begin{pmatrix} \mu_1 \\ \nu_1 \end{pmatrix}, \ R_0\eta_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.
$$

Note: $$\begin{pmatrix} \mu_1 \\ \nu_1 \end{pmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} \mu_1 \\ \nu_1 \end{pmatrix}.$$

- Assume
$$\delta \sim N(0, \kappa I_q).$$
Consider the initial condition
$$\begin{aligned} a_1 &= a \\ P_1 &= \kappa P_\infty + P_* \end{aligned} \tag{20}$$
where $P_\infty = AA'$ (from $A\delta$) and $P_* = R_0 Q_0 R_0'$ (from $R_0 \eta_0$). P_∞ is a diagonal matrix with q diagonal elements equal to one and the others equal to zero.
- We will let $\kappa \to \infty$ at a later stage (**diffuse initialization**). The Kalman filter that we derive as $\kappa \to \infty$ is called the **exact initial Kalman filter**.

- Analogously to (20), we have

$$P_t = \kappa P_{\infty,t} + P_{*,t}. \tag{21}$$

- It will be shown that $P_{\infty,t} = 0$ for $t > d$ (a positive integer), so that the usual Kalman filter applies without change for $t = d + 1, \ldots, n$ with $P_t = P_{*,t}$.
- When all initial state elements have a known joint distribution or are fixed and known, $P_\infty = 0$ and therefore $d = 0$.

- Under (21), we have

$$
\begin{aligned}
F_t \ (&= Z_t P_t Z_t' + H_t) = \kappa F_{\infty,t} + F_{*,t}, \\
M_t \ (&= P_t Z_t') = \kappa M_{\infty,t} + M_{*,t},
\end{aligned}
$$

where

$$
\begin{aligned}
F_{\infty,t} &= Z_t P_{\infty,t} Z_t', \;\; F_{*,t} = Z_t P_{*,t} Z_t' + H_t, \\
M_{\infty,t} &= P_{\infty,t} Z_t', \;\; M_{*,t} = P_{*,t} Z_t'.
\end{aligned}
$$

- Assume F_t is nonsingular. Write (cf. Koopman, S. J. (1997). Exact initial Kalman filtering and smoothing for nonstationary time series models. Journal of the American Statistical Association, 92(440), 1630-1638.)

$$F_t^{-1} = F_t^{(0)} + \kappa^{-1} F_t^{(1)} + \kappa^{-2} F_t^{(2)} + O(\kappa^{-3}).$$

Then

$$\begin{aligned} I_p &= (\kappa F_{\infty,t} + F_{*,t}) \\ &\quad \times \left(F_t^{(0)} + \kappa^{-1} F_t^{(1)} + \kappa^{-2} F_t^{(2)} + \ldots \right). \end{aligned}$$

- On equating coefficients of κ^j for $j = 1, 0, -1$, we obtain

$$\begin{aligned} F_{\infty,t} F_t^{(0)} &= 0 \ (j = 1); \ F_{\infty,t} F_t^{(1)} + F_{*,t} F_t^{(0)} = I_p \ (j = 0) \\ F_{*,t} F_t^{(1)} + F_{\infty,t} F_t^{(2)} &= 0 \ (j = -1). \end{aligned}$$

- Assume $F_{\infty,t}$ is nonsingular. Then,

$$\begin{aligned} F_t^{(0)} &= 0, \\ F_t^{(1)} &= F_{\infty,t}^{-1}, \\ F_t^{(2)} &= -F_{\infty,t}^{-1} F_{*,t} F_t^{(1)}. \end{aligned}$$

- For $K_t = T_t M_t F_t^{-1}$ and $L_t = T_t - K_t Z_t$, we have

$$K_t = T_t \left(\kappa M_{\infty,t} + M_{*,t}\right) \left(\kappa^{-1} F_t^{(1)} + \kappa^{-2} F_t^{(2)} + \ldots\right), \tag{22}$$

which gives

$$K_t = K_t^{(0)} + \kappa^{-1} K_t^{(1)} + O(\kappa^{-2}),\ L_t = L_t^{(0)} + \kappa^{-1} L_t^{(1)} + O(\kappa^{-2}),$$

where

$$\begin{aligned} K_t^{(0)} &= T_t M_{\infty,t} F_t^{(1)},\ K_t^{(1)} = T_t M_{*,t} F_t^{(1)} + T_t M_{\infty,t} F_t^{(2)} \\ L_t^{(0)} &= T_t - K_t^{(0)} Z_t,\ L_t^{(1)} = -K_t^{(1)} Z_t. \end{aligned}$$

- Using the Kalman filter and (22), we obtain

$$a_t = a_t^{(0)} + \kappa^{-1} a_t^{(1)} + O(\kappa^{-2}),$$

where $a_1^{(0)} = a$ and $a_1^{(1)} = 0$, and

$$v_t = v_t^{(0)} + \kappa^{-1} v_t^{(1)} + O(\kappa^{-2}),$$

where $v_t^{(0)} = y_t - Z_t a_t^{(0)}$ and $v_t^{(1)} = -Z_t a_t^{(1)}$.

- The updating equation for a_{t+1} can now be expressed as

$$\begin{aligned} a_{t+1} &= T_t\left(a_t^{(0)} + \kappa^{-1}a_t^{(1)} + O(\kappa^{-2})\right) \\ &\quad + \left(K_t^{(0)} + \kappa^{-1}K_t^{(1)} + O(\kappa^{-2})\right)\left(v_t^{(0)} + \kappa^{-1}v_t^{(1)} + O(\kappa^{-2})\right), \end{aligned}$$

which becomes as $\kappa \to \infty$

$$a_{t+1}^{(0)} = T_t a_t^{(0)} + K_t^{(0)} v_t^{(0)} \tag{23}$$

with $a_1^{(0)} = a$.

- The updating equation for P_{t+1} is

$$\begin{aligned} P_{t+1} &= T_t P_t (T_t - K_t Z_t)' + R_t Q_t R_t' \\ &= T_t \left(\kappa P_{\infty,t} + P_{*,t} \right) \\ &\quad \times \left(L_t^{(0)} + \kappa^{-1} L_t^{(1)} + O(\kappa^{-2}) \right)' + R_t Q_t R_t' \end{aligned}$$

which becomes as $\kappa \longrightarrow \infty$

$$\begin{aligned} P_{\infty,t+1} &= T_t P_{\infty,t} L_t^{(0)\prime}, \; (P_{\infty,1} = AA') \qquad (24) \\ P_{*,t+1} &= T_t P_{\infty,t} L_t^{(1)\prime} + T_t P_{*,t} L_t^{(0)\prime} + R_t Q_t R_t', \; (P_{*,1} = R_0 Q_0 R_0'). \end{aligned}$$

- Equations (23) and (24) constitute the Kalman filter.
- See DK for the case $F_{\infty,t} = 0$.
- For $t > d$, $P_{\infty,t} = 0$. Thus, the usual Kalman filter can be used for $t > d$. (See DK and Koopman (1997)).

Example

Consider

$$
\begin{aligned}
y_t &= \alpha_t + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \\
\alpha_{t+1} &= \alpha_t + \eta_t, \eta_t \sim N(0, \sigma_\eta^2),
\end{aligned}
$$

for which

$$
a = 0,\ A = 1,\ \delta \sim N(0, \kappa),\ R_0\eta_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.
$$

Example

(Continued) Since

$$\begin{aligned} a_1^{(0)} &= 0,\ v_1^{(0)} = y_1,\ M_{\infty,1} = P_{\infty,1} = 1,\ F_1^{(1)} = F_{\infty,1}^{-1} = P_{\infty,1}^{-1} = 1, \\ K_1^{(0)} &= M_{\infty,1} F_1^{(1)} = 1, \end{aligned}$$

we have $a_2^{(0)} = a_1^{(0)} - K_1^{(0)} v_1^{(0)} = y_1$. In addition,

$$\begin{aligned} M_{*,1} &= P_{*,1} = 0, \\ F_1^{(1)} &= 1,\ F_1^{(2)} = -F_{*,1} = -\sigma_\varepsilon^2 \\ L_1^{(1)} &= -K_1^{(1)} = -\left(M_{*,1} F_1^{(1)} + M_{\infty,1} F_1^{(2)} \right) = \sigma_\varepsilon^2 \\ L_1^{(0)} &= 1 - K_1^{(0)} = 0, \end{aligned}$$

Example

(Continued) which give

$$
\begin{aligned}
P_{\infty,2} &= P_{\infty,1} L_1^{(0)} = 0, \\
P_{*,2} &= P_{\infty,1} L_1^{(1)} + P_{*,1} L_1^{(0)} + \sigma_\eta^2 = \sigma_\varepsilon^2 + \sigma_\eta^2.
\end{aligned}
$$

Thus, the usual Kalman filter can be used for $t = 2, 3, \ldots$

Maximum likelihood estimation

- Assume $N(a_1, P_1)$ for the initial variable α_1, where a_1 and P_1 are known. The log-likelihood function is

$$\begin{aligned} \log L(Y_n) &= \sum_{t=1}^{n} \log p(y_t \mid Y_{t-1}) \\ &= c - \frac{1}{2} \sum_{t=1}^{n} (\ln |F_t| + v_t' F_t^{-1} v_t). \end{aligned}$$

  The MLE of the unknown parameters are obtained by maximizing this function.