State Space Models

A *state space model* consists of a *state equation*,

$$\mathbf{Y}_t = \mathbf{\Phi}_t \mathbf{Y}_{t-1} + \boldsymbol{\nu}_t + \mathbf{a}_t,$$

and an *observation equation*,

$$\mathbf{z}_t = \mathbf{H}_t \mathbf{Y}_t + \boldsymbol{\mu}_t + \mathbf{b}_t,$$

where $\mathbf{Y}_t$ is a state vector with a transition matrix $\mathbf{\Phi}_t$, $\mathbf{a}_t$ are independent shocks with covariance matrices $\mathbf{A}_t$, $\mathbf{H}_t$ is the observation matrix, and $\mathbf{b}_t$ are another set of shocks with covariance matrices $\mathbf{B}_t$, which are independent of $\mathbf{a}_t$. The arrays $\mathbf{\Phi}_t$, $\boldsymbol{\nu}_t$, $\mathbf{H}_t$, and $\boldsymbol{\mu}_t$ are deterministic, and often independent of t.

Through the construction of the state vector $\mathbf{Y}_t$, the AR(1) state equation is capable of representing higher order structures. The state space representation of a time series model is not unique.

ARIMA(p,d,q) in State Space Form

Consider an ARIMA(p,d,q) process in the generalized ARMA form $\varphi(B)z_t = \theta(B)a_t$. Let $m = \max(p + d, q + 1)$, one has

$$z_t = \varphi_1 z_{t-1} + \cdots + \varphi_m z_{t-m} + a_t - \theta_1 a_{t-1} - \cdots - \theta_{m-1} a_{t-m+1}.$$

Let $y_t^{(m)} = \varphi_m z_{t-1} - \theta_{m-1} a_t$, and $y_t^{(j)} = \varphi_j z_{t-1} + y_{t-1}^{(j+1)} - \theta_{j-1} a_t$, $j < m$ ($\theta_0 = -1$). The state vector $\mathbf{Y}_t = (y_t^{(1)}, \ldots, y_t^{(m)})^T$ satisfies

$$\mathbf{Y}_t = \begin{pmatrix} \varphi_1 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{m-1} & 0 & \cdots & 1 \\ \varphi_m & 0 & \cdots & 0 \end{pmatrix} \mathbf{Y}_{t-1} + \begin{pmatrix} 1 \\ -\theta_1 \\ \vdots \\ -\theta_{m-1} \end{pmatrix} a_t.$$

It is easy to check that $z_t = y_t^{(1)}$, so the observation equation is simply $z_t = (1, 0, \ldots, 0)\mathbf{Y}_t$.

Another State Space Form for ARIMA

Recall the complementary function

$$C_t(l) = z_{t+l} - \sum_{j=0}^{l-1} \psi_j a_{t+l-j} = C_{t-1}(l+1) + \psi_l a_t.$$

Set $\mathbf{Y}_t = (C_t(0), \ldots, C_t(m-1))^T$, one has

$$\mathbf{Y}_t = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ \varphi_m & \varphi_{m-1} & \cdots & \varphi_1 \end{pmatrix} \mathbf{Y}_{t-1} + \begin{pmatrix} 1 \\ \psi_1 \\ \vdots \\ \psi_{m-1} \end{pmatrix} a_t,$$

where the last equation follows from $\varphi(B)C_{t-1}(m) = 0$. The observation equation is again $z_t = (1, 0, \ldots, 0)\mathbf{Y}_t$, as $C_t(0) = z_t$.

Remember that the ψ weights are determined from φ_j, θ_j via

$$\psi_j = \varphi_1 \psi_{j-1} + \cdots + \varphi_m \psi_{j-m} - \theta_j, \quad j > 0,$$

Example: ARIMA(1,1,1)

Consider $(1-\phi B)(1-B)z_t = (1-\theta B)a_t$ with $\varphi_1 = 1+\phi$, $\varphi_2 = -\phi$, $\theta_1 = \theta$, and $\psi_1 = 1+\phi-\theta$.

With the first representation, $Y_t^{(2)} = -\phi z_{t-1} - \theta a_t$,

$$\mathbf{Y}_t = \begin{pmatrix} 1+\phi & 1 \\ -\phi & 0 \end{pmatrix} \mathbf{Y}_{t-1} + \begin{pmatrix} 1 \\ -\theta \end{pmatrix} a_t.$$

With the second representation, $Y_t^{(2)} = (1+\phi)z_t - \phi z_{t-1} - \theta a_t$,

$$\tilde{\mathbf{Y}}_t = \begin{pmatrix} 0 & 1 \\ -\phi & 1+\phi \end{pmatrix} \tilde{\mathbf{Y}}_{t-1} + \begin{pmatrix} 1 \\ 1+\phi-\theta \end{pmatrix} a_t.$$

It is seen that $\tilde{\mathbf{Y}}_t = \begin{pmatrix} 1 & 0 \\ 1+\phi & 1 \end{pmatrix}\mathbf{Y}_t$, as $\begin{pmatrix} 1 & 0 \\ 1+\phi & 1 \end{pmatrix}\begin{pmatrix} 1 \\ -\theta \end{pmatrix} = \begin{pmatrix} 1 \\ 1+\phi-\theta \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ 1+\phi & 1 \end{pmatrix}\begin{pmatrix} 1+\phi & 1 \\ -\phi & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\phi & 1+\phi \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1+\phi & 1 \end{pmatrix}$.

Kalman Filter: Derivation

Consider Gaussian process with initial state $\mathbf{Y}_0 \sim N(\mathbf{y}_0, \mathbf{V}_0)$. The state at time 1, $\mathbf{Y}_1 = \mathbf{\Phi}_1\mathbf{Y}_0 + \boldsymbol{\nu}_1 + \mathbf{a}_1$, has mean and covariance

$$\mathbf{y}_{1|0} = \mathbf{\Phi}_1\mathbf{y}_0 + \boldsymbol{\nu}_1, \quad \mathbf{V}_{1|0} = \mathbf{\Phi}_1\mathbf{V}_0\mathbf{\Phi}_1^T + \mathbf{A}_1.$$

The joint distribution of $(\mathbf{Y}_1^T, \mathbf{z}_1^T)^T$ has mean and covariance

$$\begin{pmatrix} \mathbf{y}_{1|0} \\ \mathbf{H}_1\mathbf{y}_{1|0} + \boldsymbol{\mu}_1 \end{pmatrix}, \quad \begin{pmatrix} \mathbf{V}_{1|0} & \mathbf{V}_{1|0}\mathbf{H}_1^T \\ \mathbf{H}_1\mathbf{V}_{1|0} & \mathbf{H}_1\mathbf{V}_{1|0}\mathbf{H}_1^T + \mathbf{B}_1 \end{pmatrix}.$$

The conditional distribution of $\mathbf{Y}_1|\mathbf{z}_1$ thus has the mean

$$\mathbf{y}_1 = \mathbf{y}_{1|0} + \mathbf{V}_{1|0}\mathbf{H}_1^T(\mathbf{H}_1\mathbf{V}_{1|0}\mathbf{H}_1^T + \mathbf{B}_1)^{-1}(\mathbf{z}_1 - \mathbf{H}_1\mathbf{y}_{1|0} - \boldsymbol{\mu}_1),$$

and the covariance

$$\mathbf{V}_1 = \mathbf{V}_{1|0} - \mathbf{V}_{1|0}\mathbf{H}_1^T(\mathbf{H}_1\mathbf{V}_{1|0}\mathbf{H}_1^T + \mathbf{B}_1)^{-1}\mathbf{H}_1\mathbf{V}_{1|0}.$$

Replacing 1 by t and 0 by $t-1$, one obtains the Kalman filter.

Kalman Filter: Prediction and Updating

At time $t-1$, the *prediction equations*

$$\mathbf{y}_{t|t-1} = \mathbf{\Phi}_t \mathbf{y}_{t-1} + \boldsymbol{\nu}_t, \quad \mathbf{V}_{t|t-1} = \mathbf{\Phi}_t \mathbf{V}_{t-1} \mathbf{\Phi}_t^T + \mathbf{A}_t,$$

give the optimal estimator of $\mathbf{Y}_t$ and its error covariance. For the prediction $\hat{\mathbf{z}}_{t|t-1} = \mathbf{H}_t \mathbf{y}_{t|t-1} + \boldsymbol{\mu}_t$ of $\mathbf{z}_t$, one has the *innovation* $\mathbf{e}_t = \mathbf{z}_t - \hat{\mathbf{z}}_{t|t-1}$ with the covariance $\mathbf{\Sigma}_t = \mathbf{H}_t \mathbf{V}_{t|t-1} \mathbf{H}_t^T + \mathbf{B}_t$.

Once $\mathbf{z}_t$ becomes available, the estimator of $\mathbf{Y}_t$ is updated through

$$\mathbf{y}_t = \mathbf{y}_{t|t-1} + \mathbf{V}_{t|t-1} \mathbf{H}_t^T \mathbf{\Sigma}_t^{-1} \mathbf{e}_t,$$

which has a smaller covariance

$$\mathbf{V}_t = \mathbf{V}_{t|t-1} - \mathbf{V}_{t|t-1} \mathbf{H}_t^T \mathbf{\Sigma}_t^{-1} \mathbf{H}_t \mathbf{V}_{t|t-1}.$$

The matrix $K_t = \mathbf{V}_{t|t-1} \mathbf{H}_t^T \mathbf{\Sigma}_t^{-1}$ is the Kalman gain matrix.

Example: ARMA(1,1)

A state space representation of an ARMA(1,1) model is given by

$$\mathbf{Y}_t = \begin{pmatrix} \phi & 1 \\ 0 & 0 \end{pmatrix} \mathbf{Y}_{t-1} + \begin{pmatrix} 1 \\ -\theta \end{pmatrix} a_t,$$

where $Y_t^{(1)} = z_t$ and $Y_t^{(2)} = -\theta a_t$. Set $\mathbf{y}_0 = \mathbf{0}$ and

$$\mathbf{V}_0 = \sigma_a^2 \begin{pmatrix} (1+\theta^2-2\phi\theta)/(1-\phi^2) & -\theta \\ -\theta & \theta^2 \end{pmatrix} = \sigma_a^2 \begin{pmatrix} 1+v_0 & -\theta \\ -\theta & \theta^2 \end{pmatrix},$$

where $v_0 = (\phi-\theta)^2/(1-\phi^2)$. The updating equations give

$$\mathbf{y}_{1|0} = \mathbf{0}, \quad \mathbf{V}_{1|0} = \mathbf{\Phi}\mathbf{V}_0\mathbf{\Phi}^T + \mathbf{A} = \mathbf{V}_0.$$

The innovation is $e_1 = z_1$ and the Kalman gain matrix is $K_1 = \mathbf{V}_{1|0}(1/\sigma_a^2(1+v_0), 0)^T = \left(\begin{smallmatrix} 1 \\ -\theta/(1+v_0) \end{smallmatrix}\right)$, so $\mathbf{y}_1 = \left(\begin{smallmatrix} z_1 \\ -\theta z_1/(1+v_0) \end{smallmatrix}\right)$ and $\mathbf{V}_1 = (I - K_1(1,0))\mathbf{V}_{1|0} = \sigma_a^2 v_1 \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, where $v_1 = \theta^2 v_0/(1+v_0)$. Note that $a_1|z_1$ is *not* degenerate.

Example: ARMA(1,1)

At time $t-1>0$, let $V_{t-1}=\sigma_a^2 v_{t-1}\left(\begin{smallmatrix}0&0\\0&1\end{smallmatrix}\right)$. The updating equations give $y^{(1)}_{t|t-1}=\phi z_{t-1}-\theta\tilde{a}_{t-1}$, where $\tilde{a}_{t-1}=E[a_{t-1}|z_{t-1},\ldots,z_1]$, $y^{(2)}_{t|t-1}=0$, and

$$\mathbf{V}_{t|t-1}=\mathbf{\Phi}\mathbf{V}_{t-1}\mathbf{\Phi}^T+\mathbf{A}=\sigma_a^2\begin{pmatrix}1+v_{t-1} & -\theta\\ -\theta & \theta^2\end{pmatrix}$$

The innovation $e_t=z_t-\phi z_{t-1}+\theta\tilde{a}_{t-1}=a_t-\theta(a_{t-1}-\tilde{a}_{t-1})$ has variance $\sigma_a^2(1+v_{t-1})$ and the Kalman gain matrix is given by $K_t=\mathbf{V}_{t|t-1}(1/\sigma_a^2(1+v_{t-1}),0)^T=\left(\begin{smallmatrix}1\\-\theta/(1+v_{t-1})\end{smallmatrix}\right)$. One has $\mathbf{y}_t=\left(\begin{smallmatrix}z_t\\-\theta e_t/(1+v_{t-1})\end{smallmatrix}\right)$ and $\mathbf{V}_t=(I-K_t(1,0))\mathbf{V}_{t|t-1}=\sigma_a^2 v_t\left(\begin{smallmatrix}0&0\\0&1\end{smallmatrix}\right)$, where $v_t=\theta^2 v_{t-1}/(1+v_{t-1})$. Note that for $|\theta|<1$, $v_t\to 0$ at an exponential rate.

Kalman Filter: Multiple Steps Ahead

To predict more than one step ahead based on information at time T, one simply bypass the updating step,

$$\mathbf{y}_{T+l|T} = \mathbf{\Phi}_{T+l}\mathbf{y}_{T+l-1|T} + \boldsymbol{\nu}_{T+l}, \quad l = 1, 2, \ldots,$$

where $\mathbf{y}_{T|T} = \mathbf{y}_T$. The covariance of the prediction error is given by

$$\mathbf{V}_{T+l|T} = \mathbf{\Phi}_{T+l}\mathbf{V}_{T+l-1|T}\mathbf{\Phi}_{T+l}^T + \mathbf{A}_{T+l}, \quad l = 1, 2, \ldots,$$

where $\mathbf{V}_{T|T} = \mathbf{V}_T$. The predictor of $\mathbf{z}_{T+l}$ is

$$\hat{\mathbf{z}}_{T+l|T} = \mathbf{H}_{T+l}\mathbf{y}_{T+l|T} + \boldsymbol{\mu}_{T+l},$$

with error covariance $\mathbf{E}_{T+l|T} = \mathbf{H}_{T+l}\mathbf{V}_{T+l|T}\mathbf{H}_{T+l}^T + \mathbf{B}_{T+l}$.

For ARMA(1,1) with $T > 0$, $\mathbf{y}_{T+l|T} = \mathbf{\Phi}^l\mathbf{y}_T = \phi^{l-1}\mathbf{y}_{T+1|T}$, and $\mathbf{V}_{T+l|T} = \sigma_a^2 v_{T+l|T} \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right) + \mathbf{A}$, where

$$v_{T+1|T} = v_T, \quad v_{T+l|T} = \phi^2 v_{T+l-1|T} + (\phi - \theta)^2 \to v_0.$$

Kalman Filter: Maximum Likelihood

Recall that the joint likelihood of $\mathbf{z}_1, \ldots, \mathbf{z}_N$ can be factored as $L(Z_N) = \prod_{t=1}^{N} p(\mathbf{z}_t | Z_{t-1})$, where $Z_{t-1} = \{\mathbf{z}_1, \ldots, \mathbf{z}_{t-1}\}$. For Gaussian processes, $p(\mathbf{z}_t | Z_{t-1})$ is normal with mean $\hat{\mathbf{z}}_{t|t-1}$ and covariance $\mathbf{\Sigma}_t$. For $\mathbf{z}_t$ univariate, drop boldface and write ς_t for $\mathbf{\Sigma}_t$, one has the *prediction error decomposition* form of the likelihood,

$$\log L(Z_N) = -\frac{N}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{N}\log\varsigma_t - \frac{1}{2}\sum_{t=1}^{N}\frac{e_t^2}{\varsigma_t}.$$

As a function of model parameters, $\log L(Z_N)$ can be maximized using optimization tools to yield the MLE of parameters.

For ARMA(1,1), one has $\varsigma_t = \sigma_a^2(1 + v_{t-1})$. Dropping the constant,

$$\log L(Z_N) = -\frac{1}{2}\sum_{t=1}^{N}\log(1 + v_{t-1}) - \frac{N}{2}\log\sigma_a^2 - \frac{1}{2\sigma_a^2}\sum_{t=1}^{N}\frac{e_t^2}{1+v_{t-1}}.$$

One can “profile” out σ_a^2, then work on the profile likelihood.

Example: ARIMA(1,1,1)

A state space representation of an ARIMA(1,1,1) model is

$$\mathbf{Y}_t = \begin{pmatrix} 1+\phi & 1 \\ -\phi & 0 \end{pmatrix} \mathbf{Y}_{t-1} + \begin{pmatrix} 1 \\ -\theta \end{pmatrix} a_t,$$

where $Y_t^{(1)} = z_t$ and $Y_t^{(2)} = -\phi z_{t-1} - \theta a_t$.

Write $\mathbf{V}_{1|0} = \left(\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right)$. $K_1 = \left(\begin{smallmatrix} 1 \\ b/a \end{smallmatrix}\right)$, $\mathbf{V}_1 = \begin{pmatrix} 0 & 0 \\ 0 & c-b^2/a \end{pmatrix} = \sigma_a^2 v_1 \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)$.

For $t > 1$, $\mathbf{V}_{t|t-1} = \mathbf{\Phi}\mathbf{V}_{t-1}\mathbf{\Phi}^T + \mathbf{A} = \sigma_a^2 \begin{pmatrix} 1+v_{t-1} & -\theta \\ -\theta & \theta^2 \end{pmatrix}$,
$K_t = \left(\begin{smallmatrix} 1 \\ -\theta/(1+v_{t-1}) \end{smallmatrix}\right)$, $\mathbf{V}_t = \sigma_a^2 v_t \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ for $v_t = \theta^2 v_{t-1}/(1+v_{t-1})$, following the same calculus as in ARMA(1,1). For multiple steps ahead, however, the filter does not simplify as in ARMA(1,1).

Obviousely, the same formulas hold for ARMA(2,1) in general.

Example: AR(2)

Consider a stationary AR(2) model in state space form,

$$\mathbf{Y}_t = \begin{pmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{pmatrix} \mathbf{Y}_{t-1} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} a_t,$$

where $Y_t^{(1)} = z_t$ and $Y_t^{(2)} = \phi_2 z_{t-1}$.

Set $\mathbf{V}_0 = \gamma_0 \left(\begin{smallmatrix} 1 & \phi_2\rho_1 \\ \phi_2\rho_1 & \phi_2^2 \end{smallmatrix} \right)$, the stationary variance-covariance of $\mathbf{Y}_t$.
Remember that $\rho_2 = \phi_1\rho_1 + \phi_2$ and $\sigma_a^2 = \gamma_0(1 - \phi_1\rho_1 - \phi_2\rho_2)$.
Some algebra yields $\mathbf{V}_{1|0} = \mathbf{\Phi}\mathbf{V}_0\mathbf{\Phi}^T + \mathbf{A} = \mathbf{V}_0$. $K_1 = \left(\begin{smallmatrix} 1 \\ \phi_2\rho_1 \end{smallmatrix} \right)$,
$\mathbf{V}_1 = \tilde{v}_1 \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix} \right)$ for $\tilde{v}_1 = \gamma_0\phi_2^2(1 - \rho_1^2)$.

$\mathbf{V}_{2|1} = \mathbf{\Phi}\mathbf{V}_1\mathbf{\Phi}^T + \mathbf{A} = \left(\begin{smallmatrix} \tilde{v}_1 + \sigma_a^2 & 0 \\ 0 & 0 \end{smallmatrix} \right)$, where $\tilde{v}_1 + \sigma_a^2 = \gamma_0(1 - \rho_1^2)$ as $\rho_1 = \phi_1 + \phi_2\rho_1$. $K_2 = (1, 0)^T$, $\mathbf{V}_2 = O$.

For $t > 2$, $\mathbf{V}_{t|t-1} = A = \left(\begin{smallmatrix} \sigma_a^2 & 0 \\ 0 & 0 \end{smallmatrix} \right)$, $K_t = (1, 0)^T$, $\mathbf{V}_t = O$.