State Space Models

Feb 2, 2012

Junhui Qian

1 The Model

A state space model is given by

$$y_t = Aw_t + Bx_t + u_t \tag{1}$$

$$w_t = Tw_{t-1} + v_t, \tag{2}$$

where y_t is a vector of observed variables, x_t is a vector of exogenous or predetermined variables, and w_t is a vector of possibly unobserved variables. Equation 1 is often referred to as the measurement equation, and (2) is often called the transition equation. The residuals u_t and v_t are write noises that are uncorrelated with each other. We assume $\mathbb{E}u_tu_t' = R$ and $\mathbb{E}v_tv_t' = Q$. And we often assume (u_t, v_t) are jointly normal.

Example 1: AR(p) Recall that the AP(p) model can be written in VAR(1) form as follows,

$$\begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{pmatrix} = \begin{pmatrix} \alpha_1 & \cdots & \alpha_{p-1} & \alpha_p \\ 1 & & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Let $w_t = (y_t, y_{t-1}, \ldots, y_{t-p+1})'$, the above equation can be regarded as the transition equation in state-space representation. The measurement equation is a naive identity, setting $A = (1, 0, \ldots, 0)$, $B = 0$, and $u_t = 0$.

Example 2: ARMA(p,q) Consider an ARMA(p,q) process, $\alpha(L)y_t = \beta(L)\varepsilon_t$. Let $z_t = \alpha(L)^{-1}\varepsilon_t$, we have

$$\begin{aligned} y_t &= \beta(L)z_t \\ \alpha(L)z_t &= \varepsilon_t \end{aligned}$$

If we let $r = \max(p, q+1)$ and define $w_t = (z_t, \ldots, z_{t-r+1})'$, we can easily write the model in state-space form. Consider ARMA$(1,1)$ for an example,

$$(1-\alpha L)y_t = (1+\beta L)\varepsilon_t.$$

We let

$$w_t = \begin{pmatrix} z_t \\ z_{t-1} \end{pmatrix}, \; T = \begin{pmatrix} \alpha & 0 \\ 1 & 0 \end{pmatrix}, \; v_t = \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}, \; A = (1 \;\beta), \; B = 0, \; u_t = 0.$$

2 Kalman Filter

The Kalman filter recursively estimates unobserved state variable (w_t) using observable albeit noisy signal (y_t). Let $\mathcal{F}_t$ be the natural filtration of (y_t) and assume that (x_t) is adapted to $(\mathcal{F}_t)$. Denote

$$w_{s|t} = \mathbb{E}(w_s|\mathcal{F}_t), \quad y_{s|t} = \mathbb{E}(y_s|\mathcal{F}_t),$$

and

$$\Omega_{s|t} = \text{var}(w_s|\mathcal{F}_t), \quad \Sigma_{s|t} = \text{var}(y_s|\mathcal{F}_t).$$

The Kalman filter operates in two steps, prediction and updating.

Prediction Note first that, under normality, $(y_s)_{s=1}^{t-1}$ are independent of u_t and v_t. To see this, we write $y_t = A(T^t w_0 + T^{t-1} v_1 + \cdots + T v_{t-1} + v_t) + B x_t + u_t$. It is clear that y_s is uncorrelated with v_t and u_t for all $s < t$. Conditioning the measurement and transition equations on $\mathcal{F}_{t-1}$, we then obtain the predictions of the unobservable w_t as well as the observable y_t.

$$\begin{aligned} w_{t|t-1} &= T w_{t-1|t-1} \\ y_{t|t-1} &= A w_{t|t-1} + B x_t. \end{aligned}$$

The conditional variances of these predictions are

$$\begin{aligned} \Omega_{t|t-1} &= T \Omega_{t-1|t-1} T' + Q \\ \Sigma_{t|t-1} &= A \Omega_{t|t-1} A' + R \end{aligned}$$

Updating This step updates our knowledge of w_t given the observation of y_t. Under normality, we have

$$\begin{pmatrix} w_t \\ y_t \end{pmatrix} \mid \mathcal{F}_{t-1} = N \left(\begin{pmatrix} w_{t|t-1} \\ y_{t|t-1} \end{pmatrix}, \begin{pmatrix} \Omega_{t|t-1} & \Omega_{t|t-1} A' \\ A \Omega_{t|t-1} & \Sigma_{t|t-1} \end{pmatrix} \right).$$

Then we have

$$\begin{aligned} w_{t|t} &= \mathbb{E}(w_t | y_t, \mathcal{F}_{t-1}) \\ &= w_{t|t-1} + \Omega_{t|t-1} A' \Sigma_{t|t-1}^{-1} (y_t - y_{t|t-1}), \end{aligned}$$

and

$$\begin{aligned} \Omega_{t|t} &= \mathrm{var}(w_t | y_t, \mathcal{F}_{t-1}) \\ &= \Omega_{t|t-1} - \Omega_{t|t-1} A' \Sigma_{t|t-1}^{-1} A \Omega_{t|t-1}. \end{aligned}$$

Note that $(w_{t|t} - w_{t|t-1})$ is proportional to the forecast error $(y_t - y_{t|t-1})$. The proportion $K_t = \Omega_{t|t-1} A' \Sigma_{t|t-1}^{-1}$ is called the Kalman gain. As the weight assigned to the new information, the Kalman gain is proportional to the conditional variance of w_t and inversely proportional to the conditional variance of y_t (signal).

MLE Unknown parameters in the state-space equations can be estimated by MLE. Under normality, we have

$$y_t|\mathcal{F}_{t-1} \sim N(y_{t|t-1}, \Sigma_{t|t-1}).$$

Then the log likelihood of $(y_1, y_2, \dots, y_t)$ is given by

$$\begin{aligned}
\mathcal{L}_t &= \sum_{s=1}^{t} \ell_s(\theta) \\
&= -\frac{t}{2}\log(2\pi) - \frac{1}{2}\sum_{s=1}^{t} \log \det \Sigma_{s|s-1} - \frac{1}{2}\sum_{s=1}^{t} (y_t - y_{t|t-1})' \Sigma_{t|t-1}^{-1} (y_t - y_{t|t-1}).
\end{aligned}$$

Both $y_{t|t-1}$ and $\Sigma_{t|t-1}$ are functions of parameters θ. They are iteratively computed from the prediction and updating steps of the Kalman filter given an initial value of $w_{0|0}$.

Smoothing Sometimes the inference of w_t given all observations of $(y_1, y_2, \dots, y_n)$ is of interest. More specifically, we may be interested in $w_{t|n} = \mathbb{E}(w_t|\mathcal{F}_n)$ and $\Omega_{t|n} = \mathrm{var}(w_t|\mathcal{F}_n)$. Under normality, since

$$y_{t+k} = A\left(T^{k-1} w_{t+1} + T^{k-2} v_{t+2} + \cdots + v_{t+k}\right) + B x_{t+k} + u_{t+k},$$

y_{t+k} is independent of w_t given w_{t+1} and $\mathcal{F}_t$ for all $k \geq 1$. Thus we have

$$\mathbb{E}(w_t|w_{t+1}, \mathcal{F}_n) = \mathbb{E}(w_t|w_{t+1}, \mathcal{F}_t).$$

At the same time, since

$$\begin{pmatrix} w_t \\ w_{t+1} \end{pmatrix} \mid \mathcal{F}_t = N\left(\begin{pmatrix} w_{t|t} \\ w_{t+1|t} \end{pmatrix}, \begin{pmatrix} \Omega_{t|t} & \Omega_{t|t}T' \\ T\Omega_{t|t} & \Omega_{t+1|t} \end{pmatrix}\right),$$

we have

$$\begin{aligned} w_{t|n} &= \mathbb{E}\left(\mathbb{E}\left(w_t|w_{t+1}, \mathcal{F}_n\right)|\mathcal{F}_n\right) \\ &= \mathbb{E}\left(\mathbb{E}\left(w_t|w_{t+1}, \mathcal{F}_t\right)|\mathcal{F}_n\right) \\ &= \mathbb{E}\left(w_{t|t} + \Omega_{t|t}T'\Omega_{t+1|t}^{-1}(w_{t+1} - w_{t+1|t})|\mathcal{F}_n\right). \\ &= w_{t|t} + J_t(w_{t+1|n} - w_{t+1|t}), \end{aligned}$$

where $J_t = \Omega_{t|t}T'\Omega_{t+1|t}^{-1}$. To obtain $\Omega_{t|n}$, we rewrite the above equation as,

$$w_{t|n} - w_{t|t} = J_t(w_{t+1|n} - w_{t+1|t}). \tag{3}$$

Note that $w_{t|t} = \mathbb{E}(w_{t|n}|\mathcal{F}_t)$, we thus have [1]

$$\mathbb{E}(w_{t|n} - w_{t|t})(w_{t|n} - w_{t|t})' = \mathbb{E}w_{t|n}^2 - \mathbb{E}w_{t|t}^2.$$

Furthermore,

$$\begin{aligned} \mathbb{E}w_{t|n}^2 &= \mathbb{E}\left(\mathbb{E}^2(w_t|\mathcal{F}_n)\right) \\ &= \mathbb{E}\left(-\Omega_{t|n} + \mathbb{E}(w_t^2|\mathcal{F}_n)\right) \\ &= -\Omega_{t|n} + \mathbb{E}w_t^2. \end{aligned}$$

[1] Here we use the following result: Suppose we have two random vectors X and Y and $Y = \mathbb{E}(X|\mathcal{F})$, we have $\mathbb{E}(X-Y)(X-Y)' = \mathbb{E}XX' - \mathbb{E}YY'$, since $\mathbb{E}XY' = \mathbb{E}[X\mathbb{E}(X|\mathcal{F})] = \mathbb{E}\left\{\mathbb{E}[X\mathbb{E}(X|\mathcal{F})]|\mathcal{F}\right\} = \mathbb{E}YY'$.

Similarly, we have

$$\mathbb{E}w_{t|t}^2 = -\Omega_{t|t} + \mathbb{E}w_t^2.$$

Hence, if we take expectation of the outer product of both sides of (3). The left-hand side gives,

$$\mathbb{E}(w_{t|n} - w_{t|t})(w_{t|n} - w_{t|t})' = \Omega_{t|t} - \Omega_{t|n}.$$

The right-hand side similarly gives,

$$J_t\mathbb{E}(w_{t+1|n} - w_{t+1|t})(w_{t+1|n} - w_{t+1|t})'J_t' = J_t(\Omega_{t+1|t} - \Omega_{t+1|n})J_t'.$$

Rearranging terms, we obtain

$$\Omega_{t|n} = \Omega_{t|t} + J_t(\Omega_{t+1|n} - \Omega_{t+1|t})J_t',$$

which can be iteratively applied to estimate all $(\Omega_{t|n})$, given $\Omega_{n|n}$ which is known in the update step.

Nonnormality If we do not assume normality, the above results are still applicable, although we need to interpret them differently. We may make the following convention,

- $\mathcal{F}_t$ linear space spanned by $(y_s)_{s=1}^t$
- $\mathbb{E}(\cdot|\mathcal{F}_t)$ projection on $\mathcal{F}_t$
- $\text{var}(\cdot|\mathcal{F}_t)$ variance of $\mathcal{F}_t$-projection error.

Under this convention, the Kalman filter finds the minimum mean square linear estimate and its mean squared error.

3 Markov Switching Autoregressive Model (Hamilton's)

Hamilton's Markov Switching Autoregressive Model (MSAR) is given by

$$\begin{aligned} y_t &= \mu_{s_t} + w_t \\ \alpha(L) w_t &= \varepsilon_t, \end{aligned}$$

where $\alpha(z) = 1 - \alpha_1 L - \cdots - \alpha_p z^p$, (s_t) is a series of unobservable state variables that is described by a Markov chain conditional on $(\mathcal{F}_{t-1})$, and ε_t is assumed to be i.i.d. $\mathrm{N}(0, \sigma^2)$.

We assume that the process s_t can take values in $\{1, \ldots, N\}$, where N is the number of states or regimes. Markov chain may be the simplest model for such a discrete-valued process. It assumes that the probability of s_t being in some state depends on the past only through the value of s_{t-1},

$$\mathbb{P}\{s_t = j | s_{t-1} = i, s_{t-2} = k, \ldots, \} = \mathbb{P}\{s_t = j | s_{t-1} = i\} \equiv p_{ij}.$$

$\{p_{ij}\}$ satisfies $\sum_{j=1}^{N} p_{ij} = 1$ for all i and are called transition probabilities. We may conveniently $\{p_{ij}\}$ in matrix form,

$$P = \begin{pmatrix} p_{11} & p_{21} & \cdots & p_{N1} \\ p_{21} & p_{22} & \cdots & p_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1} & p_{N1} & \cdots & p_{NN} \end{pmatrix}.$$

P is called the transition matrix. Using the transition matrix, we may represent a Markov chain with a vector autoregression. First, let $(e_i, i = 1, \ldots, N)$ be $N \times 1$ vectors with 1 on the i-th element and 0's on the remaining. We define a vector-valued process ξ_t,

$$\xi_t = \begin{cases} e_1 = (1, 0, 0, \ldots, 0)' & \text{when } s_t = 1 \\ e_2 = (0, 1, 0, \ldots, 0)' & \text{when } s_t = 2 \\ \vdots & \vdots \\ e_N = (0, 0, 0, \ldots, 1)' & \text{when } s_t = N \end{cases}$$

Note that

$$\mathbb{E}(\xi_t | \xi_{t-1} = e_i) = \mathbb{E}(\xi_t | s_{t-1} = i) = \begin{pmatrix} p_{i1} \\ p_{i2} \\ \vdots \\ p_{iN} \end{pmatrix} = Pe_i.$$

We thus have

$$\mathbb{E}(\xi_t | \xi_{t-1}) = P\xi_{t-1}.$$

Furthermore, it is straightforward to show that

$$\mathbb{E}(\xi_{t+m} | \xi_t) = P^m \xi_t.$$

Example: A two-state MSAR(1) model The model is defined as follows,

$$\begin{aligned} y_t &= \mu_{s_t} + w_t, \ s_t \in \{1, 2\} \\ w_t &= \alpha w_{t-1} + \varepsilon_t \\ P &= \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}. \end{aligned}$$

The parameters in this model include μ_1, μ_2, α, σ^2, p, and q.

MLE We first consider MSAR(1) models. Our results extend easily to more general MSAR(p) models. To obtain likelihood for $(y_1, \ldots, y_n)$, we iteratively calculate $p(y_t | \mathcal{F}_{t-1})$.

In each step, the calculation has two steps, prediction and updating, much like the Kalman filter. The prediction step is

$$p(y_t|\mathcal{F}_{t-1}) = \sum_{s_t, s_{t-1}} p(y_t|s_t, s_{t-1}, \mathcal{F}_{t-1})p(s_t, s_{t-1}|\mathcal{F}_{t-1}),$$

where

$$p(s_t, s_{t-1}|\mathcal{F}_{t-1}) = p(s_t|s_{t-1})p(s_{t-1}|\mathcal{F}_{t-1}).$$

Note that $p(s_t|s_{t-1})$ is transition probabilities. The updating step calculates $p(s_t, s_{t-1}|\mathcal{F}_t)$, marginal integration of which gives $p(s_t|\mathcal{F}_t)$,

$$p(s_t, s_{t-1}|\mathcal{F}_t) = \frac{p(y_t|s_t, s_{t-1}, \mathcal{F}_{t-1})p(s_t, s_{t-1}|\mathcal{F}_{t-1})}{p(y_t|\mathcal{F}_{t-1})}.$$

For more general MSAR(p), we consider $p(s_t, s_{t-1}, \ldots, s_{t-p}|\mathcal{F}_{t-1})$ and $p(s_t, s_{t-1}, \ldots, s_{t-p}|\mathcal{F}_t)$ in place of $p(s_t, s_{t-1}|\mathcal{F}_{t-1})$ and $p(s_t, s_{t-1}|\mathcal{F}_t)$, respectively.

Smoothing We may make more precise statements on the state/regime at time t using full-sample, like the smoothing step in the Kalman filter. Note first that we already have $p(s_n, s_{n-1}|\mathcal{F}_n)$, hence it suffices to find a way to calculate $p(s_t, s_{t-1}|\mathcal{F}_n)$ given $p(s_{t+1}, s_t|\mathcal{F}_n)$. It is well known from the Bayes formula that

$$p(x|y, z) = p(x|z) \quad \text{if} \quad p(y|x, z) = p(y|z).$$

Since $p(y_{t+k}|s_{t+1}, s_t, s_{t-1}, \mathcal{F}_t) = p(y_{t+k}|s_{t+1}, s_t, \mathcal{F}_t)$ for all $k \geq 1$, we have

$$p(s_{t-1}|s_{t+1}, s_t, \mathcal{F}_n) = p(s_{t-1}|s_{t+1}, s_t, \mathcal{F}_t).$$

Similarly, since $p(s_{t+1}|s_t, s_{t-1}, \mathcal{F}_t) = p(s_{t+1}|s_t, \mathcal{F}_t)$, we have

$$p(s_{t-1}|s_{t+1}, s_t, \mathcal{F}_t) = p(s_{t-1}|s_t, \mathcal{F}_t) = \frac{p(s_t, s_{t-1}|\mathcal{F}_t)}{p(s_t|\mathcal{F}_t)}.$$

Now we can calculate

$$p(s_{t+1}, s_t, s_{t-1}|\mathcal{F}_n) = p(s_{t-1}|s_{t+1}, s_t, \mathcal{F}_n)p(s_{t+1}, s_t|\mathcal{F}_n).$$

Taking marginal integration (summation) of $p(s_{t+1}, s_t, s_{t-1}|\mathcal{F}_n)$ along the first dimension, we obtain $p(s_t, s_{t-1}|\mathcal{F}_n)$. Similarly, the smoothing step of MSAR(p) models calculates $p(s_t, \ldots, s_{t-p}|\mathcal{F}_n)$ from $p(s_{t+1}, \ldots, s_{t-p+1}|\mathcal{F}_n)$. We have

$$p(s_{t+1}, \ldots, s_{t-p}|\mathcal{F}_n) = p(s_{t-p}|s_{t+1}, \ldots, s_{t-p+1}, \mathcal{F}_n)p(s_{t+1}, \ldots, s_{t-p}|\mathcal{F}_n),$$

where

$$\begin{aligned} p(s_{t-p}|s_{t+1}, \ldots, s_{t-p+1}, \mathcal{F}_n) &= p(s_{t-p}|s_{t+1}, \ldots, s_{t-p+1}, \mathcal{F}_t) \\ &= p(s_{t-p}|s_t, \ldots, s_{t-p+1}, \mathcal{F}_t) \\ &= \frac{p(s_t, \ldots, s_{t-p+1}, s_{t-p}|\mathcal{F}_t)}{p(s_t, \ldots, s_{t-p+1}|\mathcal{F}_t)}. \end{aligned}$$

References

Hamilton, J.D., 1989, A new approach to the economic analysis of nonstationary time series and the business cycle, Econometrica, 57, 357-384.