14.384 Time Series Analysis, Fall 2007
Professor Anna Mikusheva
Paul Schrimpf, scribe
Novemeber 15, 2007
revised November 24, 2009

Lecture 21

Filtering. State space models. Kalman filter.

State-Space Models

In this lecture we consider state-space models, which often appear in macro, as well as other areas of economics.

Example 1. For example, suppose, GDP growth, y_t is given by

$$\begin{aligned} y_t =& \mu_t + \epsilon_t \\ \mu_t =& \mu_{t-1} + \eta_t \end{aligned}$$

where μ_t is the slow moving component of GDP growth and ϵ_t is noise with $(\epsilon_t, \eta_t) \sim iid\, N\left(0, \begin{pmatrix} \sigma_\epsilon^2 & 0 \\ 0 & \sigma_\eta^2 \end{pmatrix}\right)$

. Assume that a researcher observe y_t, but not the latent trend μ_t.

Example 2. *Markov Switching*

$$\begin{aligned} y_t =& \beta_0 + \beta_1 S_t + \epsilon_t \\ S_t \in& \{0, 1\} \\ P(S_t = 1 | S_{t-1} = 0) =& 1 - q \\ P(S_t = 1 | S_{t-1} = 1) =& p \end{aligned}$$

If y_t is GDP growth, we might think of S_t as representing whether or not we're in a boom. We observe y_t, but not the latent state S_t.

What is common for the two examples above is that there are some observed variable y_t whose behavior depends on unobserved state (μ_t or S_t), let's call it α_t. Some questions we might want to answer in these examples include:

1. Estimate parameters: e.g. in example 1 estimate σ_ϵ and σ_η
2. Extract unobserved state: e.g. in example 1 estimate μ_t
3. Forecast future values of y_t.

In a state space model, we have an (potentially unobserved) state variable, α_t, and measurements, y_t. Let $\mathcal{Y}_{t-1}$ be all measurable ($\{y_1, ..., y_{t-1}\}$) variables up to time $t-1$. The state space model is characterized by

(1) State equation, describing the evolution of state:

$$F(\alpha_t | \alpha_{t-1}, \mathcal{Y}_{t-1})$$

(2) Measurement equation, describing how the measurable variables relate to state variables

$$f(y_t | \alpha_t, \mathcal{Y}_{t-1})$$

The two equations allow us to write the joint likelihood of observed variables $(y_1, y_2, ..., y_T)$. There could be some unknown parameter θ which we will drop for simplicity later on.

$$f(y_1, ..., y_T; \theta) = f(y_1; \theta) \prod_{t=2}^{T} f(y_t | y_{t-1}, ..., y_1; \theta) = f(y_1; \theta) \prod_{t=2}^{T} f(y_t | \mathcal{Y}_{t-1}; \theta)$$

For the likelihood, we need to find $f(y_t|\mathcal{Y}_{t-1})$. This can be done in three general steps and is called filtering:

1. $$f(y_t|\mathcal{Y}_{t-1}) = \int f(y_t|\alpha_t, \mathcal{Y}_{t-1}) f(\alpha_t|\mathcal{Y}_{t-1}) d\alpha_t$$

2. Prediction equation

$$f(\alpha_t|\mathcal{Y}_{t-1}) = \int F(\alpha_t|\alpha_{t-1}, \mathcal{Y}_{t-1}) f(\alpha_{t-1}|\mathcal{Y}_{t-1}) d\alpha_{t-1}$$

3. Updating equation

$$f(\alpha_t|\mathcal{Y}_t) = \frac{f(y_t|\alpha_t, \mathcal{Y}_{t-1}) f(\alpha_t|\mathcal{Y}_{t-1})}{f(y_t|\mathcal{Y}_{t-1})}$$

On a theoretical level, this process is clear and straightforward. We go from $f(\alpha_1|\mathcal{Y}_0)$ to $f(y_1|\mathcal{Y}_0)$ to $f(\alpha_1|\mathcal{Y}_1)$ to $f(\alpha_2|\mathcal{Y}_1)$ to finally get $f(y_2|y_1)$, the conditional likelihood. However, notice that the two first steps include integration, and in practice, it is usually difficult to compute these integrals. There are two cases where the integration is straightforward. With discrete distributions, the integrals are just sums. With normal distributions, sub-vector of normal vector is normal, and conditionals are normal as well, this simplify life a lot. The corresponding procedure for normals is called Kalman filter. For other situations, integration is very difficult, one approach is called particle filtering.

Kalman Filtering

Suppose we have a state model:

$$\alpha_t = T\alpha_{t-1} + R\eta_t \tag{1}$$

and a measurement:

$$y_t = Z\alpha_t + S\xi_t \tag{2}$$

with $\begin{pmatrix} \eta_t \\ \xi_t \end{pmatrix} \sim iid\, N\left(0, \begin{pmatrix} Q & 0 \\ 0 & H \end{pmatrix}\right)$. Then,

$$\begin{aligned} F(\alpha_t|\alpha_{t-1}) &\sim N(T\alpha_{t-1}, RQR') \\ f(y_t|\alpha_t, \mathcal{Y}_{t-1}) &\sim N(Z\alpha_t, SHS') \end{aligned}$$

If α_1 is normal, then since α_t's and y_t's are linear combinations of normal errors, the vector, $(\alpha_1, ..., \alpha_T, y_1, ..., y_T)$ is normally distributed. We will use the general fact that if

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$$

then

$$x_1|x_2 \sim N(\tilde{\mu}, \tilde{\Sigma})$$

with $\tilde{\mu} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$ and $\tilde{\Sigma} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$

Using this result, we see that each of the distributions in the general steps above are normal. Since the normal distribution is characterized by mean and variance, we need only compute them. Let us introduce the following notation:

$$\begin{aligned} \alpha_t|\mathcal{Y}_{t-1} \sim & N(\alpha_{t|t-1}, P_{t|t-1}) \\ \alpha_t|\mathcal{Y}_t \sim & N(\alpha_{t|t}, P_{t|t}) \\ y_t|\mathcal{Y}_{t-1} \sim & N(y_{t|t-1}, F_t) \end{aligned}$$

From equation (1), we see that

$$\alpha_{t|t-1} = T\alpha_{t-1|t-1} \tag{3}$$

$$P_{t|t-1} = E\left((\alpha_t - \alpha_{t|t-1})(\alpha_t - \alpha_{t|t-1})'|\mathcal{Y}_{t-1}\right) = TP_{t-1|t-1}T' + RQR' \tag{4}$$

Here we used that $\alpha_t - \alpha_{t|t-1} = T(\alpha_{t-1} - \alpha_{t-1|t-1}) + R\eta_t$. These two equations can be used in step 2 above (prediction equation).

Now, looking at (2), we'll get the equations that we need for step 1.

$$y_{t|t-1} = Z\alpha_{t|t-1} \tag{5}$$

$$F_t = E\left((y_t - y_{t|t-1})(y_t - y_{t|t-1})'|\mathcal{Y}_{t-1}\right) = ZP_{t|t-1}Z' + SHS' \tag{6}$$

Note that so far we have only used the linearity of of (1) and (2).

For the updating step 3, we will need to use normality.

$$\begin{pmatrix} \alpha_t \\ y_t \end{pmatrix} |\mathcal{Y}_{t-1} \sim N\left(\begin{pmatrix} \alpha_{t|t-1} \\ y_{t|t-1} \end{pmatrix}, \begin{pmatrix} P_{t|t-1} & ? \\ ? & F_t \end{pmatrix} \right)$$

where $? = E((\alpha_t - \alpha_{t|t-1})(y_t - y_{t-1|t})'|\mathcal{Y}_{t-1}) = E((\alpha_t - \alpha_{t|t-1})(\alpha_t - \alpha_{t-1|t})'Z'|\mathcal{Y}_{t-1}) = P_{t|t-1}Z'$. We can use this and the general fact about normals to write the posteriority density of α_t given $\mathcal{Y}_t$.

$$\begin{aligned} \alpha_t|\mathcal{Y}_t = \alpha_t|(y_t, \mathcal{Y}_{t-1}) \sim & N(\alpha_{t|t}, P_{t|t}) \\ \sim & N(\alpha_{t|t-1} + P_{t|t-1}Z'F_t^{-1}(y_t - y_{t|t-1}), P_{t|t-1} - P_{t|t-1}Z'F_t^{-1}ZP_{t|t-1}) \end{aligned} \tag{7}$$

So, starting from some initial $\alpha_{1|0}$ and $P_{1|0}$ we use (5) and (6) to get $y_{1|0}$ and F_1 (the conditional density of y_1). Then using (7), we can get $\alpha_{1|1}$ and $P_{1|1}$. From there, we use (3) and (4) to get $\alpha_{2|1}$ and $P_{2|1}$. Repeating in this way, we can compute the entire likelihood (conditional on the initial conditions). We could just go ahead and use this procedure for computing the likelihood and then estimate the parameters by MLE.

Kalman Smoother

The Kalman filter uses data on the past and current observations, $\mathcal{Y}_t$, to predict α_t. This is what we want for computing the likelihood. However, you might want to estimate α_t. For this, you want to use all the data to predict α_t. This is called the Kalman smoother. The idea is as follows: let

$$E(\alpha_t|\mathcal{Y}_T) = \alpha_{t|T}$$

We know that $(\alpha_t, \alpha_{t+1})|\mathcal{Y}_t$ is normal, so

$$\begin{aligned} E(\alpha_t|\alpha_{t+1}, \mathcal{Y}_t) = & \alpha_{t|t} + E\left((\alpha_t - \alpha_{t|t})(\alpha_{t+1} - \alpha_{t+1|t})'|\mathcal{Y}_t\right) P_{t+1|t+1}^{-1}(\alpha_{t+1} - \alpha_{t+1|t}) \\ = & \alpha_{t|t} + J_t(\alpha_{t+1} - \alpha_{t+1|t}) \end{aligned}$$

given that $E\left((\alpha_t - \alpha_{t|t})(\alpha_{t+1} - \alpha_{t+1|t})'|\mathcal{Y}_t\right) = E\left((\alpha_t - \alpha_{t|t})(T(\alpha_t - \alpha_{t|t})) + R\eta_{t+1}|\mathcal{Y}_t\right) = P_{t|t}T$ we have $J_t = P_{t|t}T'P_{t+1|t}^{-1}$. So then,

$$E(\alpha_t|\alpha_{t+1}, \mathcal{Y}_t) = \alpha_{t|t} + J_t(\alpha_{t+1} - \alpha_{t+1|t})$$

Expectation $E(\alpha_t|\alpha_{t+1}, \mathcal{Y}_T)$ is the same as $E(\alpha_t|\alpha_{t+1}, \mathcal{Y}_t)$, because the knowledge of y_{t+j} for $j > 0$ would be of no added value if we already knew α_{t+1}. That is,

$$E(\alpha_t|\alpha_{t+1}, \mathcal{Y}_T) = \alpha_{t|t} + J_t(\alpha_{t+1} - \alpha_{t+1|t})$$

No we project the results on $\mathcal{Y}_T$ only and use the law of iterated expectations:

$$E(\alpha_t|\mathcal{Y}_T) = \alpha_{t|t} + J_t(\alpha_{t+1|T} - \alpha_{t+1|t})$$

Starting from $t = T$ and repeating in this way, we can compute $\alpha_{T|T}$, $\alpha_{T-1|T}$, ..., $\alpha_{1|T}$.

Things to remember: the Kalman filter and smoother are linear in data. The Kalman filter is a recursive procedure running forward. After that, we can run the Kalman smoother backward.

Summary

For a state-space model,

$$y_t = Z_t\alpha_t + S_t\xi_t$$
$$\alpha_t = T_t\alpha_{t-1} + R_t\eta_t$$

with $\begin{pmatrix} \eta_t \\ \xi_t \end{pmatrix} \sim iid\, N\left(0, \begin{bmatrix} Q & 0 \\ 0 & H \end{bmatrix}\right)$. and the initial observation $y_1 \sim N(y_{1|0}, F_1)$. Kalman filter can be used for the following tasks:

- construct the conditional distribution of y_t and α_t given all past observations $|\mathcal{Y}_{t-1}$. We know that it is normal, Kalman filter calculates the conditional mean $y_{t|t-1}, \alpha_{t|t-1}$ and conditional variances.
- calculate a likelihood of $\{y_1, ..., y_T\}$
- construct the distribution of α given all observations of y, $\alpha_t|\mathcal{Y}_T$ (extract trend- Kalman smoother).
- calculate $y_{t|t-1}$, which is the best linear forecast of y_t given $|\mathcal{Y}_{t-1}$ even if errors are non-normal

What can be cast into state-space form?

We can write many models in state-space form:

Example 3. $AR(p)$:

$$y_t = \sum_{i=1}^{P} \phi_i y_{t-1} + \epsilon_t$$

Then our state equation is

$$\alpha_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \dots & \phi_p \\ 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix} \alpha_{t-1} + \begin{bmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

The measurement equation is

$$y_t = [1\,0\,...\,0]\alpha_t$$

For an AR model, it is straightforward to write down the likelihood directly, so there is no need to write down the state-space form or use the Kalman filter. However, for MA and $ARMA$ models, the likelihood is very difficult to evaluate without using the Kalman filter.

Example 4. *MA* Model:

$$y_t = \epsilon_t + \theta\epsilon_{t-1}$$

State equation:

$$\alpha_t = \begin{bmatrix} \epsilon_t \\ \epsilon_{t-1} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \alpha_{t-1} + \begin{bmatrix} \epsilon_t \\ 0 \end{bmatrix}$$

Measurement:

$$y_t = [1\,\theta]\alpha_t$$

Example 5. *ARMA* Model:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t + \theta\epsilon_{t-1}$$

State equation:

$$\alpha_t = \begin{bmatrix} y_t \\ \phi_2 y_{t-1} + \theta\epsilon_t \end{bmatrix} = \begin{bmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ \phi_2 y_{t-2} + \theta\epsilon_{t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ \theta \end{bmatrix} \epsilon_t$$

Measurement:

$$y_t = [1\,0]\alpha_t$$

What else can be done by Kalman filter? Error-in-variables

State-space models can also be used to analyze error-in-variable models. Assume that we want to model the behavior the ex ante real interest rate, ξ_t as (for example) an autoregressive process

$$\xi_t = \phi\xi_{t-1} + u_t$$

and want to estimate ϕ and make future forecasts. The problem though is that the theoretical ex ante real interest rate

$$\xi_t = i_t - \pi_t^e$$

is not observable since the expected inflation, π_t^e, is not observable. Instead we observe the realized interest rate, $i_t - \pi_t$ (and the realized inflation). The problem can be written in a state space model with measurement equation:

$$\xi_t^{realized} = i_t - \pi_t = \xi_t - (\pi_t - \pi_t^e) = \xi_t + v_t$$

and state equation:

$$\xi_t = \phi\xi_{t-1} + u_t$$

Since $v_t = \pi_t - \pi_t^e$ is the prediction error, it's natural to assume the iid structure of it.

Missing or unequally spaced Observations

Suppose we have a state-space model where we are missing observations. The model is:

$$\begin{aligned} y_t &= z_t\alpha_t + S_t\xi_t \\ \alpha_t &= T_t\alpha_{t-1} + R_t\eta_t \end{aligned}$$

But instead of observing all $\{y_t\}_{t=1}^T$, we only observe some subset $\{y_{i1}, .., y_{i\tau}\} = \{y_i | i \in I\}$. To produce a likelihood we can consider the following model:

$$\begin{aligned} y_t^* &= z_t^*\alpha_t + w_t^* \\ \alpha_t &= T_t\alpha_{t-1} + R_t\eta_t \end{aligned}$$

where $y_t^* = \begin{cases} y_t & t \in I \\ N_t & \text{otherwise} \end{cases}$, $z_t^* = \begin{cases} z_t & t \in I \\ 0 & \text{otherwise} \end{cases}$, and $w_t^* = \begin{cases} w_t & t \in I \\ N_t & \text{otherwise} \end{cases}$, with $N_t \sim iid\, N(0,1)$. Assume that we observe $y_t^* = 0$ when $t \notin I$, then the conditional likelihood is

$$f(y_t^*|\mathcal{Y}_{t-1}^*) = \begin{cases} \phi(0) & t \notin I \\ f(y_t|\{y_s : s \in I,\, s < t\}) & \end{cases}$$

and the likelihood is $f(y_1^*, ..., y_T^*) = \phi(0)^{\#I^c} f(y_{i1}, ..., y_{i\tau}) = (2\pi)^{-\frac{\#I_c}{2}} f(y_{i1}, ..., y_{i\tau})$. It means that we can run Kalman filter for y_t^*, it produces(up to a constant) a valid likelihood for the initial model.

Example 6. Suppose:

$$y_t = \phi y_{t-1} + \epsilon_t$$

and we observe $t \in I = \{1, 3, 4, 5\}$. In state space form, as above, we have:

$$\begin{aligned} \alpha_t &= \alpha_{t-1}\phi + \epsilon_t \\ y_t^* &= z_t \alpha_t + w_t \end{aligned}$$

where $z_t = \begin{cases} 1 & t \in I \\ 0 & \text{otherwise} \end{cases}$ and $w_t = \begin{cases} 0 & t \in I \\ N(0,1) & \text{otherwise} \end{cases}$. Let's think what Kalman filter would do: for $t \in I$ we observe $y_t = \alpha_t$, so our best linear predictor of α_t is y_t. For $t = 2$, y_2 is unrelated to α_2, so our best linear predictor of α_2 is just ϕy_1. Then the conditional means used in the Kalman filter are:

$$\alpha_{t|t} = \begin{cases} y_t & t \in I \\ \phi y_{t-1} & t = 2 \end{cases}$$

To form the conditional likelihood, we need the distribution of $y_t|\mathcal{Y}_{t-1}$, which has mean

$$y_{t|t-1}^* = \phi \alpha_{t-1|t-1} = \begin{cases} \phi y_{t-1} & t \in \{4, 5\} \\ \phi^2 y_1 & t = 3 \end{cases}$$

and variance

$$F_t = \begin{cases} \sigma^2 & t \in \{4, 5\} \\ (1 + \phi^2)\sigma^2 & t = 3 \end{cases}$$

We only need the conditional distribution at $t = 3, 4, 5$ because the likelihood is:

$$f(y_1, y_3, y_4, y_5) = f(y_1) f(y_3|y_1) f(y_4|y_3, y_1) f(y_5|y_4, y_3, y_1)$$

and the conditional (on y_1) likelihood is

$$\begin{aligned} f(y_1, y_3, y_4, y_5|y_1) =& f(y_3|y_1) f(y_4|y_3, y_1) f(y_5|y_4, y_3, y_1) \\ =& \frac{1}{\sigma\sqrt{1+\phi^2}} \phi\left(\frac{y_3 - \phi^2 y_1}{\sigma\sqrt{1+\phi^2}}\right) \frac{1}{\sigma}\phi\left(\frac{y_4 - \phi y_3}{\sigma}\right) \frac{1}{\sigma}\phi\left(\frac{y_5 - \phi y_4}{\sigma}\right) \end{aligned}$$

where $\phi(\cdot)$ is the normal pdf.

The nice thing is that Kalman filter does this reasoning automatically.