14.384 Time Series Analysis
Fall 2008

14.384 Time Series Analysis, Fall 2007
Professor Anna Mikusheva
Paul Schrimpf, scribe
Novemeber 15, 2007

Lecture 20

Filtering

State-Space Models

The Kalman filter is widely used to compute state-space models. These 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 $(\epsilon_t, \eta_t) \sim iid\, N\left(0, \begin{pmatrix} \sigma_\epsilon^2 & 0 \\ 0 & \sigma_\eta^2 \end{pmatrix}\right)$

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.

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 trend: e.g. in example 1 estimate μ_t
3. Forecasting

We can estimate the parameters by maximum likelihood. Often, it useful to write the joint likelihood of $(y_1, y_2, ..., y_T)$ as a product of conditional densities,

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

In a state space model, we have an unobserved state variable, α_t, and measurements, y_t. The state variables are distributed according to a state equation,

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

where $\mathcal{Y}_{t-1}$ are all measurable $(\{y_1, ..., y_{t-1}\})$ variables up to time $t-1$. and the measurable variables have a measurement equation,

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

For the likelihood, we need $f(y_t | \mathcal{Y}_{t-1})$. We can compute this by integrating out α_t. The general steps are:

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, in practice, it is usually difficult to compute these integrals. There are two cases where the integration is straightforward. With normal distributions, we can use Kalman filtering. With discrete distributions, the integrals are just sums. For other situations, it 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

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

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

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 $? = P_{t|t-1}Z'$. 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. However, there are a few loose ends that we have not dealt with. There are unknown initial conditions that we haven't talked about. Also, our data is not iid, so the usual results about the consistency and efficiency of MLE do not apply. We will talk about these issues next time.

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

$$\begin{aligned} E(\alpha_t|\alpha_{t+1}, \mathcal{Y}_t) &= \alpha_{t|t} + \text{cov}(\alpha_t, \alpha_{t+1})P^{-1}_{t+1|t+1}(\alpha_{t+1} - \alpha_{t+1|t}) \\ &= \alpha_{t|t} + J_t(\alpha_{t+1} - \alpha_{t+1|t}) \end{aligned}$$

where $J_t = P_{t|t}T'P^{-1}_{t+1|t}$ so then,

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

where we used the fact that $\mathcal{Y}_T$ is richer than $\mathcal{Y}_t$. From this, we receive:

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

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.