Class notes: Advanced Topics in Macroeconomics

Topic: Kalman Filter

Date: October 1, 2018

Once we have learned to compute equilibria of near-linear models, we want to estimate their parameters. To do that, we'll rely on a standard filtering technique called the Kalman filter and we'll derive the likelihood function given a set of observables. For most of our problems, we need the Kalman filter because we do not observe the entire state space. We'll start off the lecture with some warm-up exercises with problems where all the data are observable. That will help us see the difference in the likelihood functions.

1. Warm-up Exercises

Let's start by estimating an AR(1) process:

$$Y_t = c + \phi Y_{t-1} + \epsilon_t$$

where $E\epsilon_t = 0$, $E\epsilon_t^2 = \sigma^2$, $E\epsilon_t \epsilon_s = 0$ if $t \neq s$. Suppose that we observe the sample $(y_1, y_2, \ldots, y_T)$ and want to estimate ϕ. We want to calculate the probability density function over the random variables $Y_1, \ldots Y_T$, say $f_{Y_1, Y_2, \ldots Y_T}(y_1, y_2, \ldots, y_T; \theta)$, which is the probability of observing the particular sample given the parameter vector **theta** $= [c, \phi, \sigma^2]'$. The maximum likelihood estimate (MLE) of θ is the value that maximizes the density. Before we can do that, we have to specify a particular distribution for ϵ_t. Suppose it is normally distributed, that is, $\epsilon_t \sim N(0, \sigma^2)$.

The probability distribution of Y_1 is

$$f_{Y_1}(y_1; \theta) = \frac{1}{\sqrt{2\pi\sigma^2/(1-\phi^2)}} \exp\left(-\{y - c/(1-\phi)\}^2 / \left(2\sigma^2 / \left(1-\phi^2\right)\right)\right)$$

since the normal density for a variable with mean μ and variance σ^2 is given by:

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-(y-\mu)^2 / \left(2\sigma^2\right)\right)$$

and the mean of Y_1 is $c/(1-\phi)$ and the variance is $\sigma^2/(1-\phi^2)$.

Given the Markovian structure of the stochastic process, the joint distribution of Y_1, Y_2 is given by the product of the marginal distributions:

$$f_{Y_1,Y_2}\left(y_1,y_2;\theta\right) = f_{Y_2|Y_1}\left(y_2|y_1;\theta\right)$$

and so is that of any subsample:

$$f_{Y_1,Y_2,\ldots Y_t}\left(y_1,y_2,\ldots,y_t;\theta\right) = f_{Y_1}\left(y_1;\theta\right)\prod_{t=2}^{T} f_{Y_t|Y_{t-1}}\left(y_t|y_{t-1};\theta\right).$$

The log-likelihood function, $\mathcal{L}(\theta)$ is the log of this density:

$$\begin{aligned}
\mathcal{L}\left(\theta\right) &= \log f_{Y_1}\left(y_1;\mathbf{theta}\right) + \sum_{t=2}^{T}\log f_{Y_t|Y_{t-1}}\left(y_t|y_{t-1};\theta\right) \\
&= -\frac{1}{2}\log\left(2\pi\right) - \frac{1}{2}\log\left(\sigma^2/\left(1-\phi^2\right)\right) \\
&\quad - \frac{\left(y_1 - c/\left(1-\phi\right)\right)^2}{2\sigma^2/\left(1-\phi^2\right)} - \left(T-1\right)/2\log\left(2\pi\right) \\
&\quad - \left(T-1\right)/2\log\sigma^2 - \sum_{t=2}^{T}\left(y_t - c - \phi y_{t-1}\right)^2/\left(2\sigma^2\right)
\end{aligned}$$

To compute the (exact) MLE estimates, we take the derivatives of $\mathcal{L}(\theta)$ with respect to each element of θ and set them equal to zero. Because θ enters f_{Y_1}, this leads to a set of nonlinear functions in θ and the sample of y_t's. An alternative procedure is to treat y_1 as deterministic and maximize the likelihood function conditional on it. Then, we are maximizing

$$\mathcal{L}\left(\theta\right) = -\left(T-1\right)/2\log\sigma^2 - \sum_{t=2}^{T}\left(y_t - c - \phi y_{t-1}\right)^2/\left(2\sigma^2\right)$$

If we maximize this with respect to c, ϕ, and σ, we get the following result familiar to those who use ordinary least squares:

$$\begin{bmatrix}\hat{c}\\ \hat{\phi}\end{bmatrix} = \begin{bmatrix}T-1 & \sum y_{t-1}\\ \sum y_{t-1} & \hat{y}_t^2\end{bmatrix}^{-1}\begin{bmatrix}\sum y_t\\ \sum y_t y_{t-1}\end{bmatrix}$$

with the summation over $t = 2, \ldots, T$, and

$$\hat{\sigma}^2 = \sum_{t=2}^{T}\left(y_t - \hat{c} - \hat{\phi}y_{t-1}\right)^2/\left(T-1\right).$$

In class, have the students demonstrate that an AR(p) process has likelihood function, conditioned on the first p observations is:

$$\mathcal{L}(\theta) = -(T-p)/2\log\sigma^2 - \sum_{t=p+1}^{T}(y_t - c - \phi_1 y_{t-1} - \ldots \phi_p y_{t-p})^2/(2\sigma^2)$$

where here the constant term is ignored. Again, the MLE estimates for c and the ϕ's are the same as the OLS estimates since we would be minimizing the sum of residuals squared and the MLE for variance is given by:

$$\hat{\sigma}^2 = \sum_{t=p+1}^{T}\left(y_t - \hat{c} - \hat{\phi}_1 y_{t-1} - \ldots \hat{\phi}_p y_{t-p}\right)^2/(T-p).$$

Next, we consider cases when we cannot observe all of the variables and have to use a model to make predictions via the Kalman filter.

2. Kalman Filter

Below, I'll assume all shocks are Gaussian and therefore I can use the following relations that hold for jointly normal random variables, say Y and X. If the mean of Y is $\bar{Y}$, the mean of X is $\bar{X}$ and the variance-covariance matrix is Σ, where

$$\Sigma = \begin{bmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{bmatrix}$$

then the conditional expectation of Y given X is

$$E[Y|X] = \bar{Y} + \Sigma_{YX}\Sigma_{XX}^{-1}\left(X - \bar{X}\right) \tag{2.1}$$

and the conditional variance is

$$\mathrm{Var}[Y|X] = E\left[(Y - E(Y|X))^2|X\right] = \Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}. \tag{2.2}$$

Below, we will use these formulas to construct means and variances of a prediction of latent state variables.

To derive the Kalman filtering equations, I will use the notation of Harvey (1989), which is a very useful textbook for time series econometricians. Assume that we have a *state space system* as follows:

$$\alpha_t = T\alpha_{t-1} + \eta_t \tag{2.3}$$

$$y_t = Z\alpha_t + \epsilon_t \tag{2.4}$$

where α and η are $m \times 1$ vectors, T is $m \times m$, y and ϵ are $n \times 1$ vectors, Z is $n \times m$, $E\eta_t = 0$, $E\eta_t\eta_t' = Q$, $E\epsilon_t = 0$, $E\epsilon_t\epsilon_t' = H$, and $E\epsilon_t\eta_s = 0$ for all t, and s. The variables in α_t are unobserved and need to be estimated and the variables in y_t are observed and will be used to construct estimates. Equation (2.3) is the transition equation for the unobserved state vector and (2.4) is the measurement or observer equation.

The ultimate goal here is to estimate parameters of the model. To do this, we will convert the system in (2.3)-(2.4) to:

$$\hat{\alpha}_{t+1|t} = T\hat{\alpha}_{t|t-1} + K_t v_t \tag{2.5}$$

$$y_t = Z\hat{\alpha}_{t|t-1} + v_t \tag{2.6}$$

where $\hat{\alpha}_{t|t-1}$ is an estimate of the unobserved state vector (defined specifically later) and v_t is an *innovation*, which is the difference between the vector of observables and a forecast of that vector. The matrix K_t is the *Kalman gain*, which will be derived below. To estimate parameters, we need a time series for the innovations and an estimate of its variance-covariance matrix. In other words, we want estimates of parameters that make these errors small. Those estimates will be most "likely" to have generated the sequence of $\{y_t\}$ that we observe.

To get the sequence of innovations, we compute a sequence of recursions. Suppose at time t that we have an estimate of

$$\hat{\alpha}_{t-1} = E\left[\alpha_{t-1}|y_0, y_1, ..., y_{t-1}\right]$$

and the estimation error for α_{t-1} is P_{t-1}:

$$P_{t-1} = E\left[\left(\alpha_{t-1} - \hat{\alpha}_{t-1}\right)\left(\alpha_{t-1} - \hat{\alpha}_{t-1}\right)'\right].$$

Given $\hat{\alpha}_{t-1}$ and P_{t-1}, we can estimate α_t using (2.3):

$$\hat{\alpha}_{t|t-1} = T\alpha_{t-1} \tag{2.7}$$

which is a function of the lagged observations and $\hat{\alpha}_{t-1}$, P_{t-1}. The variance of the prediction error is

$$\begin{aligned}
P_{t|t-1} &= E\left[\left(\alpha_t - \hat{\alpha}_{t|t-1}\right)\left(\alpha_t - \hat{\alpha}_{t|t-1}\right)'\right] \\
&= E\left[\left(T\alpha_{t-1} + \eta_t - T\hat{\alpha}_{t-1}\right)\left(T\alpha_{t-1} + \eta_t - T\hat{\alpha}_{t-1}\right)'\right] \\
&= E\left[\left(T\left(\alpha_{t-1} - \hat{\alpha}_{t-1}\right) + \eta_t\right)\left(T\left(\alpha_{t-1} - \hat{\alpha}_{t-1}\right) + \eta_t\right)'\right] \\
&= TE\left[\left(\alpha_{t-1} - \hat{\alpha}_{t-1}\right)\left(\alpha_{t-1} - \hat{\alpha}_{t-1}\right)'\right]T' + E\eta_t\eta_t' \\
&= TP_{t-1}T' + Q,
\end{aligned} \tag{2.8}$$

which follows from the fact that η_t is not correlated with $\alpha_{t-1} - \hat{\alpha}_{t-1}$.

Next, consider updating these equations after observing y_t. In other words, we update $E[\alpha_t|y_0, y_1, ..., y_{t-1}]$ to $E[\alpha_t|y_0, y_1, ..., y_t]$. The best estimate for y_t given $\alpha_{t|t-1}$ is

$$\hat{y}_{t|t-1} = Z\hat{\alpha}_{t|t-1}$$

with the estimation error being the innovation we seek:

$$v_t = y_t - \hat{y}_{t|t-1} = Z\left(\alpha_t - \hat{\alpha}_{t|t-1}\right) + \epsilon_t. \tag{2.9}$$

The variance of this innovation is given by

$$\begin{aligned}
F_t &= Ev_tv_t' \\
&= E\left[\left(Z\left(\alpha_t - \hat{\alpha}_{t|t-1}\right) + \epsilon_t\right)\left(Z\left(\alpha_t - \hat{\alpha}_{t|t-1}\right) + \epsilon_t\right)'\right] \\
&= ZE\left[\left(\alpha_t - \hat{\alpha}_{t|t-1}\right)\left(\alpha_t - \hat{\alpha}_{t|t-1}\right)'\right]Z' + E\epsilon_t\epsilon_t' \\
&= ZP_{t|t-1}Z' + H
\end{aligned} \tag{2.10}$$

and the covariance of v_t with the estimation error is:

$$\begin{aligned}
G_t &= E\left[\left(y_t - \hat{y}_{t|t-1}\right)\left(\alpha_t - \hat{\alpha}_{t|t-1}\right)'\right] \\
&= E\left[\left(Z\left(\alpha_t - \hat{\alpha}_{t|t-1}\right) + \epsilon_t\right)\left(\alpha_t - \hat{\alpha}_{t|t-1}\right)'\right] \\
&= ZE\left[\left(\alpha_t - \hat{\alpha}_{t|t-1}\right)\left(\alpha_t - \hat{\alpha}_{t|t-1}\right)'\right] \\
&= ZP_{t|t-1}.
\end{aligned}$$

We now have everything that we need to apply the formulas in (2.1) and (2.2). Assume that α is like Y and y is like X, then the conditional mean is

$$\underbrace{\hat{\alpha}_t}_{E[Y|X]} = \underbrace{\hat{\alpha}_{t|t-1}}_{EY} + \underbrace{P_{t|t-1}Z'}_{\Sigma_{YX}} \left(\underbrace{ZP_{t|t-1}Z' + H}_{\Sigma_{XX}=F_t} \right)^{-1} \left(y_t - \underbrace{Z\alpha_{t|t-1}}_{EX} \right) \tag{2.11}$$

and the conditional variance is

$$\underbrace{P_t}_{\mathrm{Var}(Y|X)} = \underbrace{P_{t|t-1}}_{\Sigma_{YY}} - \underbrace{P_{t|t-1}Z'}_{\Sigma_{YX}} \left(\underbrace{ZP_{t|t-1}Z' + H}_{\Sigma_{XX}} \right)^{-1} \underbrace{ZP_{t|t-1}}_{\Sigma_{XY}} \tag{2.12}$$

The idea here is that the new information in y_t—the observed prediction error—is used to update the estimate of α_t from $\hat{\alpha}_{t|t-1}$ to $\hat{\alpha}_t$.

Multiplying both sides of (2.11) by T, we get

$$\begin{aligned} \hat{\alpha}_{t+1|t} &= T\hat{\alpha}_{t|t-1} + TP_{t|t-1}Z' \left(ZP_{t|t-1}Z' + H \right)^{-1} v_t \\ &= T\hat{\alpha}_{t|t-1} + K_t v_t \end{aligned}$$

where K_t can also be written more succinctly as

$$K_t = TP_{t|t-1}Z'F_t^{-1}.$$

The Kalman algorithm can now be implemented. Starting with guesses for α_0 and P_0, recursively update the estimates of the mean and variance of the state using (2.7), (2.8), (2.11) and (2.12) (in that order). Along the way, store v_t and F_t using (2.9) and (2.10). In stationary models, one can replace P_0 with the stationary P that solves $P = TPT' + Q$ and set α_0 to the unconditional mean of the state vector. To compute parameters, we need to maximize the log-likelihood function:

$$\ln L = \sum_t \{ -\frac{n}{2} \ln 2\pi - \frac{1}{2} \ln |F_t| - \frac{1}{2} v_t' F_t^{-1} v_t \}.$$