Econ 8185: Quant PS3

Bipul Verma

December 26, 2021

Abstract

The present document uses Kalman Filter to estimate the coefficients and shock variance.

1 State Space Models and Kalman Filter

The linear state space model assumes that observed time series $\{y_t\}$ is a linear function of state variable α_t (usually unobserved). It is further assumed that the state variable α_t follows a first order vector autoregressive process. Mathematically the *state space* system is:

$$y_t = Z\alpha_t + \epsilon_t \tag{1}$$
$$\alpha_t = T\alpha_{t-1} + \eta_t \tag{2}$$

where ϵ_t, η_t are assumed to be mean 0 white noise. Specifically,

$$\mathbb{E}[\epsilon_t] = 0, \mathbb{E}[\eta_t] = 0, \mathbb{E}[\epsilon_t \eta_s] = 0$$
$$\mathbb{E}[\epsilon_t \epsilon_t'] = H$$
$$\mathbb{E}[\eta_t \eta_t'] = Q.$$

Equation (1) is also called as the *Measurement or Observer equation*, while equation (2) as the *transition equation.* It turns out that a lot of models can be put in state space representation. Once we have the state space representation, the Kalman Filter is a system of two equations- the *predicting equation* and the *updating equation.* Mathematically, the Kalman Filter consists of:

$$\hat{\alpha}_{t|t-1} = T\hat{\alpha}_{t-1} \tag{3}$$
$$\hat{\alpha}_{t+1|t} = T\hat{\alpha}_{t|t-1} + K_t i_t \tag{4}$$
$$y_t = \underbrace{Z\hat{\alpha}_{t|t-1}}_{\hat{y}_{t|t-1}} + i_t \tag{5}$$

where $\hat{\alpha}_{t|t-1}$ is an estimate of unobserved state vector. The matrix K_t is *Kalman gain*, $i_t = y_t - \hat{y}_{t|t-1} = y_t - Z\hat{\alpha}_{t|t-1}$ is an *innovation* which is the difference between the observable and its prediction. The idea being that the actual observed y_t consists of a component that we were able to predict $\hat{y}_{t|t-1}$ and *innovation* i_t. It is to note that the prediction of y_t i.e $\hat{y}_{t|t-1}$ is entirely based on the prediction of state $\hat{\alpha}_{t|t-1}$ (since $\hat{y}_{t|t-1} = Z\hat{\alpha}_{t|t-1}$). Kalman filter thus evolves around predicting ($\hat{\alpha}_{t|t-1}$) and updating the prediction of the state vector($\hat{\alpha}_{t+1|t}$). At time t before we have actually observed y_t, our information set consists of lagged values of $y : (y_0, y_1, \ldots, y_{t-1})$. Based on this information, we begin by making the best guess for the underlying state. From the law of motion of state we have:

$$\hat{\alpha}_{t|t-1} = \mathbb{E}[\alpha_t|(y_0, y_1, \ldots, y_{t-1})] = T\mathbb{E}[\alpha_{t-1}|(y_0, y_1, \ldots, y_{t-1})] + \underbrace{\mathbb{E}[\eta_t|(y_0, y_1, \ldots, y_{t-1})]}_{0}.$$

Define:

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

Combining the above two equation gives us back equation (3), which is our prediction for the state vector, which then gives the prediction for observable $\hat{y}_{t|t-1}$. Now we want to update our prediction in after we have actually observed y_t. But before that we'll define some objects which will be used in our updation equation.

At time t, we can write the variance of prediction error as:

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

We see that our estimate $\hat{\alpha}_{t|t-1} = T\hat{\alpha}_{t-1}$ is a function of lagged values of y. The conditional variance of prediction error is given by:

$$\begin{aligned}
P_{t|t-1} &= \mathbb{E}[(\alpha_t - \hat{\alpha}_{t|t-1})(\alpha_t - \hat{\alpha}_{t|t-1})'] \\
&= \mathbb{E}[(T\alpha_{t-1} + \eta_t - T\hat{\alpha}_{t-1})(T\alpha_{t-1} + \eta_t - T\hat{\alpha}_{t-1})'] \\
&= \mathbb{E}[(T(\alpha_{t-1} - \hat{\alpha}_{t-1}) + \eta_t)(T(\alpha_{t-1} - \hat{\alpha}_{t-1}) + \eta_t)'] \\
&= T\mathbb{E}[(\alpha_{t-1} - \hat{\alpha}_{t-1})(\alpha_{t-1} - \hat{\alpha}_{t-1})']T' + \mathbb{E}[\eta_t\eta_t'] \\
&= TP_{t-1}T' + Q.
\end{aligned}$$

Here we have made use of the fact that $(\alpha_{t-1} - \hat{\alpha}_{t-1})$ and η_t are uncorrelated. In a similar spirit we can calculate the variance of innovation $i_t = y_t - y_{t|\hat{t}-1} = y_t - Z\hat{\alpha}_{t|t-1}]$. The conditional variance of innovation is:

$$\begin{aligned}
F_t &= \mathbb{E}[i_t i_t'] \\
&= \mathbb{E}[(y_t - Z\hat{\alpha}_{t|t-1})(y_t - Z\hat{\alpha}_{t|t-1})'] \\
&= \mathbb{E}[(Z\alpha_t - Z\hat{\alpha}_{t|t-1})(Z\alpha_t - Z\hat{\alpha}_{t|t-1})'] \\
&= ZP_{t|t-1}Z' + H.
\end{aligned}$$

Similarly, the covariance of i_t with prediction (state) error is:

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

Note that before we actually observe y_t, conditional on the information set $(y_0, y_1, \ldots, y_{t-1})$, all the matrices $P_{t-1}, P_{t|t-1}, F_t, G_t$, are known.

For jointly Normal random variable X and Y with the var-covariance matrix given by:

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

the conditional expectation of Y given X is given by:

$$\mathbb{E}[Y|X] = \mathbb{E}[Y] + \Sigma_{XY}\Sigma_{XX}^{-1}(X - \mathbb{E}[X]),$$

and the conditional variance is given by:

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

It then follows from the above formula that:

$$\begin{aligned}
\mathbb{E}[\alpha_t|(y_0, y_1, \ldots, y_{t-1}), y_t] = \mathbb{E}[\alpha_t|(y_0, y_1, \ldots, y_{t-1})]& \\
+ Cov(\alpha_t, y_t)Var(y_t)^{-1}(y_t - \mathbb{E}[y_t|(y_0, y_1, \ldots, y_{t-1})]).&
\end{aligned}$$

$$
\begin{aligned}
Cov(\alpha_t, y_t) &= \mathbb{E}[(\alpha_t - E[\alpha_t|(y_0, y_1, \ldots, y_{t-1})])(y_t - E[y_t|(y_0, y_1, \ldots, y_{t-1})])'] \\
&= \mathbb{E}[(\alpha_t - \hat{\alpha}_{t|t-1})(y_t - \hat{y}_{t|t-1})'] \\
&= P_{t|t-1}Z' \\
Var(y_t) &= \mathbb{E}[(y_t - E[y_t|(y_0, y_1, \ldots, y_{t-1})])(y_t - E[y_t|(y_0, y_1, \ldots, y_{t-1})']] \\
&= \mathbb{E}[(y_t - \hat{y}_{t|t-1})(y_t - \hat{y}_{t|t-1})'] \\
&= ZP_{t|t-1}Z' + H.
\end{aligned}
$$

Thus the updation equation is given by:

$$
\begin{aligned}
\hat{\alpha}_t &= \hat{\alpha}_{t|t-1} + P_{t|t-1}Z'(ZP_{t|t-1}Z' + H)^{-1}(y_t - \hat{y}_{t|t-1}) \\
&= \hat{\alpha}_{t|t-1} + G_t'(F_t)^{-1}(y_t - \hat{y}_{t|t-1}).
\end{aligned}
$$

Multiplying both sides of previous equation by T we get:

$$
\hat{\alpha}_{t+1|t} = T\hat{\alpha}_{t|t-1} + Ki_t
$$

where $K = TP_{t|t-1}Z'(ZP_{t|t-1}Z' + H)^{-1} = TG_t'(F_t)^{-1}$, and $i_t = y_t - \hat{y}_{t|t-1}$. In a similar fashion using the conditional variance formula we get:

$$
P_t = P_{t|t-1} - G_t'F_t^{-1}G_t.
$$

1.1 Kalman Filter Implementation

1. Begin with a guess for α_0, P_0. One can set α_0 to be the unconditional mean of state vector, P_0 with the stationary P that solves $P = TPT' + Q$. Set $\hat{\alpha}_{1|0} = T\alpha_0, P_{1|0} = TP_0T' + Q$.
2. Get:
   (a) $\hat{y}_{1|0} = Z\hat{\alpha}_{1|0}$
   (b) $i_1 = y_1 - \hat{y}_{1|0}$
   (c) $F_1 = ZP_{1|0}Z' + H, G_1 = ZP_{1|0}, K_1 = TG_1'F_1^{-1}$
   (d) $\hat{\alpha}_{2|1} = T\hat{\alpha}_{1|0} + K_1 i_1$
   (e) $P_{2|1} = T(P_{1|0} - G_1'F_1^{-1}G_1)T' + Q$
   (f) Go back to step (a) increasing the time order by 1 unit.
3. Continue the recursion in this fashion to get sequence of $\{\hat{\alpha}_{t|t-1}\}$, $\{i_t\}$ and $\{\hat{y}_{t|t-1}\}$.

1.2 Parametric Estimation using Log Likelihood

To estimate the parameters we set the likelihood as:

$$
\ln L = \sum_t \{-\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|det(F_t)| - \frac{1}{2}i_t'F_t^{-1}i_t
$$

1.3 Application: HW3

We'll write the processes in state space representation first.

1. $AR(1) : x_t = \rho x_{t-1} + \epsilon_t$

   $$\begin{aligned} &\text{State:} && x_t = [\rho]x_{t-1} + \epsilon_t \\ &\text{Observation:} && x_t = [1]x_t \end{aligned}$$

2. $AR(2) : x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + \epsilon_t$

   $$\text{State:} \quad \begin{bmatrix} x_t \\ x_{t-1} \end{bmatrix} = \begin{bmatrix} \rho_1 & \rho_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ x_{t-2} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \epsilon_t$$

   $$\text{Observation:} \quad x_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_t \\ x_{t-1} \end{bmatrix}$$

3. $MA(1) : x_t = \epsilon_t + \rho \epsilon_{t-1}$

   $$\text{State:} \quad \begin{bmatrix} \epsilon_t \\ \epsilon_{t-1} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \epsilon_{t-1} \\ \epsilon_{t-2} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \epsilon_t$$

   $$\text{Observation:} \quad x_t = \begin{bmatrix} 1 & \rho \end{bmatrix} \begin{bmatrix} \epsilon_t \\ \epsilon_{t-1} \end{bmatrix}$$

4. $RandomWalk :$

   $$\begin{aligned} &\text{State:} && \mu_t = [1]\mu_{t-1} + \eta_t \\ &\text{Observation:} && x_t = [1]\mu_t + \epsilon_t. \end{aligned}$$

1.4 Results

1.4.1 AR(1)

We set $\rho = 0.6$ and $\sigma = 0.5$. We then simulate values of the $AR(1)$ process for 100 periods.

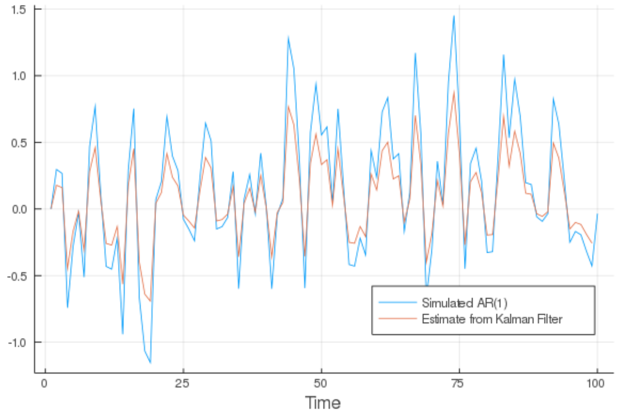

Figure 1: Simulated vs Filtered AR(1)

We note that given the required input matrices Kalman Filter closely follows the actual simulated $AR(1)$ process. We then use the log likelihood formula as given in section 1.2 to calculate the likelihood of observing the simulated $AR(1)$ process. We then maximize the likelihood with respect to the $AR(1)$ parameters to get estimates for the parameters. The estimated parameters are: $\hat{\rho} = 0.44, \hat{\sigma} = 0.48$ with 100 observation and $\hat{\rho} = 0.59, \hat{\sigma} = 0.48$ with 1000 observations. We note that the estimates are very close to original with 1000 data points.

1.4.2 AR(2)

We set $\rho_1 = 0.3, \rho_2 = 0.4$, and $\sigma = 0.5$. We then simulate values for the $AR(2)$ process for 100 periods.

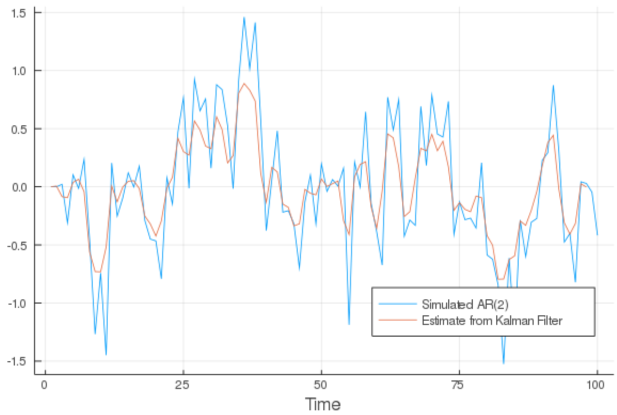

Figure 2: Simulated vs Filtered AR(2)

We note that the filtered process matches the simulated process well, except for instances where there are large spikes in the simulated process. The estimates from MLE estimates are: $\hat{\rho_1} = 0.43, \hat{\rho_2} = 0.34, \hat{\sigma} = 0.49$ with 100 observation and $\hat{\rho_1} = 0.399, \hat{\rho_2} = 0.499, \hat{\sigma} = 0.507$ with 1000 observations. We note that the estimates are very close to original with 1000 data points.

1.4.3 MA(1)

We set the parameters $\rho = 0.7, \sigma = 0.8$ for $MA(1)$ process. We simulate the values for $MA(1)$ process for 100 periods.

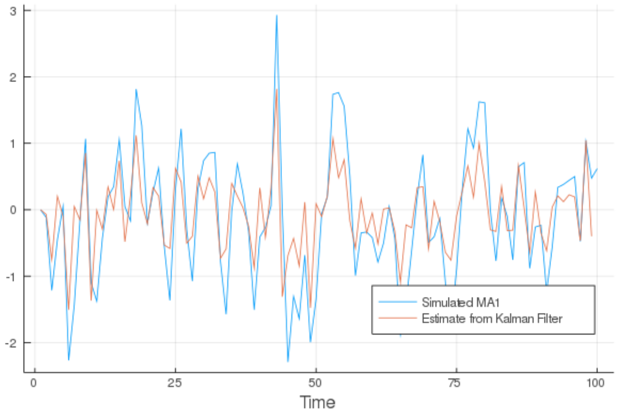

Figure 3: Simulated vs Filtered MA(1)

We note that the filtered process matches the simulated process well, except for instances where there are large spikes in the simulated process. The estimates from MLE estimates are: $\hat{\rho} = 0.79, \hat{\sigma} = 0.83$ with 100 data points, and $\hat{\rho} = 0.694, \hat{\sigma} = 0.812$ with 5000 observations . We note that the MLE estimates are close to the actual parameter values used for simulation with 5000 observations.

1.4.4 Random Walk

We set the parameters $\sigma_\epsilon = 0.8, \sigma_\eta = 0.6$ for Random Walk process. We simulate the values for Random Walk for 100 periods.

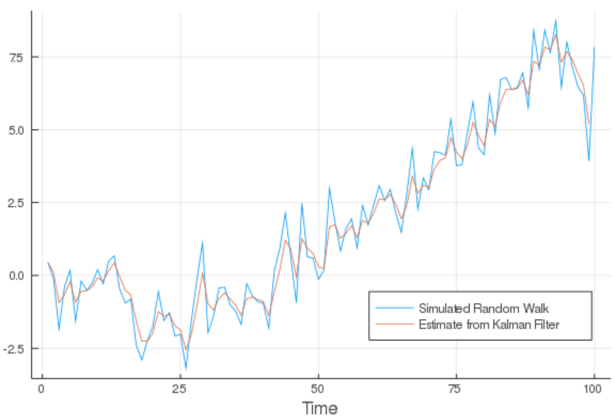

Figure 4: Simulated vs Filtered Random Walk

We note that the filtered process matches the simulated process well, except for instances where there are large spikes in the simulated process. The estimates from MLE estimates are: $\hat{\sigma}_\epsilon = 0.85, \hat{\sigma}_\eta = 0.41$ with 100 observation, and $\hat{\sigma}_\epsilon = 0.81, \hat{\sigma}_\eta = 0.58$ with 5000 observations. We note that the MLE estimates are close to the actual parameter values with 5000 observations.