Kalman Filter

H. Cetin [*]

March 14, 2023

Abstract

We explain what Kalman Filter is and apply Kalman Filter on AR(1), AR(2), MA(1), and Random Walk processes to estimate the coefficients and shock variances via maximum likelihood. The code can be found bluehere.

[*]University of Minnesota, E-mail: cetin019@umn.edu

1 The Theory

Let's assume the law of motions of the unobservable and observable RV's are the following Time series:

$$\text{Unobservable: } \alpha_t = T\alpha_{t-1} + \eta_t, \text{where, } \eta_{t+1} \sim N(0, Q)$$

$$\text{Observable: } y_t = Z\alpha_t + \epsilon_t, \text{where, } \epsilon_t \sim N(0, H)$$

Assume the shocks for both series are serially uncorrelated.

For computation purposes let's write down the dimensions of each variable. (for each observation t)

Assume there are **n** observable random variables and **m** unobservable random variables for each time t. Then the dimensions are:

- T: (m,m) - Z: (n,m) - H: (n,n) - Q: (m,m) - α_t: (m,1) - η_t: (m,1) - ϵ_t: (n,1)

Our goal is to estimate parameters (T, Q, Z, and H) by using the observable sequence.

To do that we'll connect our system:

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

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

Here:

- $\hat{\alpha}_{t+1|t}$: Estimate of the unobserved state vector,
- v_t : Innovation (Difference between the observable and the forecast)
- K_t : Kalmar gain

Our goal is to find parameters such that the error is small enough. Then, we can conclude that those parameters will be most "likely" to have generated the sequence of observables that we observed.

How do we get v_t? To get it, we need to solve a sequence of recursions:

Suppose at time t, we have:

1. **estimate**: $\hat{\alpha}_{t-1} = \mathbb{E}[\alpha_{t-1}|y_0, y_1, \dots y_{t-1}]$
2. **estimation error for** α_{t-1}: $P_{t-1} = \mathbb{E}[(\alpha_{t-1} - \hat{\alpha}_{t-1})(\alpha_{t-1} - \hat{\alpha}_{t-1})']$

Then, given $\hat{\alpha}_{t-1}, P_{t-1}$, we can estimate α_t by using **one step prediction mean and variances** as:

$$\text{(Prediciton mean): } \hat{\alpha}_{t|t-1} = T\hat{\alpha} \tag{1}$$

$$\hat{P}_{t|t-1} = \mathbb{E}[(\alpha_t - \hat{\alpha}_{t|t-1})(\alpha_t - \hat{\alpha}_{t|t-1})']$$

$$\mathbb{E}[(T\alpha_{t-1} + \eta_t - T\hat{\alpha}_{t-1})(T\alpha_{t-1} + \eta_t - T\hat{\alpha}_{t-1})']$$

$$\text{(Prediciton variance): } \hat{P}_{t|t-1} = TP_{t-1}T' + Q \tag{2}$$

Next, we update our estimate after the observation y_t: from $\mathbb{E}[\alpha_t|y_0, y_1, y_2 \dots y_{t-1}]$ to $\mathbb{E}[\alpha_t|y_0, y_1, y_2 \dots y_t]$

So far, without knowing the y_t, the best estimate for y_t is:

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

Then, define the innovation (measurement error of the observable):

$$v_t = y_t - \hat{y}_{t|t-1} = Z(\alpha_t - \hat{\alpha}_{t|t-1}) +_t \tag{3}$$

And define the variance of this innovation as:

$$Var(v_t) = F_t = \mathbb{E}[v_t v_t']$$

$$\mathbb{E}[(Z(\alpha_t - \hat{\alpha}_{t|t-1}) +_t)(Z(\alpha_t - \hat{\alpha}_{t|t-1}) +_t)']$$

$$F_t = ZTP_{t-1}T'Z' + ZQZ' + H \tag{4}$$

Then, the covariance of innovation $v_t = (y_t - \hat{y}_{t|t-1})$ (which is the measurement error of the observable) and the estimation error $(\alpha_t - \hat{\alpha}_{t|t-1})$ is:

$$G_t = \mathbb{E}[v_t(\alpha_t - \hat{\alpha}_{t|t-1})']$$

$$= \mathbb{E}[(y_t - \hat{y}_{t|t-1})(\alpha_t - \hat{\alpha}_{t|t-1})']$$

$$= \mathbb{E}[(Z(\alpha_t - \hat{\alpha}_{t|t-1}) + \epsilon_t))(\alpha_t - \hat{\alpha}_{t|t-1})']$$

$$G_t = ZP_{t|t-1} \tag{5}$$

Then, the conditional mean and variance are: (Using Gaussian RV properties)

$$\hat{\alpha}_t = \hat{\alpha}_{t|t-1} + P_{t|t-1}Z'(ZP_{t|t-1}Z' + H)^{-1}(y_t - Z\alpha_{t|t-1}) \tag{6}$$

$$P_t = P_{t|t-1} - P_{t|t-1}Z'(ZP_{t|t-1}Z' + H)^{-1}ZP_{t|t-1} \tag{7}$$

The idea here is that the new information y_t is used to estimate new prediction density with mean $\hat{\alpha}_t$ and variance P_t.

Then, we will find our new estimate for α_{t+1} by multiplying both sides of conditional mean with T:

$$\hat{\alpha}_{t+1|t} = T\hat{\alpha}_t = T\hat{\alpha}_{t|t-1} + TP_{t|t-1}Z'(ZP_{t|t-1}Z' + H)^{-1}(y_t - Z\alpha_{t|t-1}) \tag{8}$$

and so on...

So the Kalman algorithm is the following:

- For all parameters Q, Z, T, and H candidates, do the following:
  - Start with the initial guess of α_0 and P_0
  - Recursively update your $\hat{\alpha}_t's$ and $P_t's$ for each observation
  - While you are updating your estimates, store v_t and correspondingly, F_t
  - To estimate the parameters T, Q, H, and Z, do Kalman filtering for a 4-dimensional grid of estimate nominees. After each filtering, Use their v_t and $F_t's$ to compute the log-likelihood of the estimation: $log(L) = \sum_t(-\frac{n}{2}log(2\pi) - \frac{1}{2}log(|F_t|) - \frac{1}{2}v_t'F_t^{-1}v_t)$
  - Pick the parameters that maximize this log-likelihood estimate.

References

[1] McGrattan, Class Notes