Financial Econometrics

Sol Yates

March 4, 2024

Contents

Lecture 6: Kalman Filter

Mon 04 Mar 09:04

[Lecture]

$$E[\sigma_{t+1}^2|\mathcal{F}_t] = w + \alpha\varepsilon_t^2 + \beta\sigma_t^2$$

If we think about the classical ARMA-GARCH framework, we have

1. returns with some conditional mean $+\varepsilon_t$ where

$$\begin{aligned} r_t &= E[r_t|\mathcal{F}_{t-1}] + \varepsilon_t \\ \varepsilon_t &= \mathcal{L}_t \cdot \sigma_t(\mathcal{F}_{t-1}) \\ \mathcal{L}_t &\sim \mathcal{N}(0,1) \end{aligned}$$

If we would like to assume that returns is driven by

$$\begin{aligned} r_t &= \varepsilon_t = \mathcal{L}_t\sigma_t \\ \sigma_t^2 &= f(\mathcal{F}_{t-1}) + q_t \\ \varepsilon_t|\mathcal{F}_{t-1} &\sim \mathcal{N}(0,1) \\ f(\varepsilon_t) &= \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right) \end{aligned}$$

Multivariate normal distribution

$$\begin{bmatrix} r \\ y \end{bmatrix} \approx N\left(\begin{pmatrix} \mu_y \\ \mu_x \end{pmatrix} \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy} & \Sigma_{yy} \end{pmatrix}\right)$$

We need the $x|y$ distribution to derive the Kalman filter, the transformation can get us certain properties

Expected value (these are population parameters, fixed values, numbers measuring fixed variance)

$$
\begin{aligned}
E[z] &= E[x] - \Sigma_{xy}\Sigma_{yy}^{-1}(E[y] - \mu_y) \\
&= E[Z'Z] \quad \text{if scalar, then} \quad E[z^2] \\
&= E\left[\left(x - \Sigma_{xy}\Sigma_{yy}^{-1}(y-\mu_y)\right)\left(x - \Sigma_{xy}\Sigma_{yy}^{-1}(y-\mu_y)\right)\right] \\
&\quad \text{algebra} \ldots \\
\mathrm{cov}(x,y) &= E[xy] - E[x]E[y] \\
&\to u_x u_y - E[xy'] = -\mathrm{cov}(x,y) = -\Sigma_{xy} \\
&\quad \text{and so the whole term} \\
&= E[X'X] - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}' \\
\mathrm{cov}(z,z) &= E[ZZ'] - E[Z]E[Z]' \\
&= E[xx']
\end{aligned}
$$

We have 2 RV with joint distribution, we want to understand the conditional distribution

$$X|y \approx N\left(\ldots\right)$$

Then if we take the transformation z

$$z = x - \Sigma_{xy}$$

1. $E[z] = E[x]$
2. $V[z] = \Sigma_{xx} - \Sigma_{yy}^{-1}\Sigma_{xy}$
3. $\mathrm{Cov}(y,z) = 0$

$x = z + \Sigma_{xy} + \Sigma_{yy}^{-1}(y - \mu y)$
$E[x|y] = E[z|y] + \Sigma_{xy}\Sigma_{yy}^{-1}(E[y|y] - \mu_y)$
First term $= E[z]$, second term is same third is y so

$$E[x|y] = \mu_x = \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y)$$

$V[x|y]$

$$x|y \approx N\left(\mu_x\right)$$

Local Trend Model

In order to understand the SV model, consider a simple local trend model first

$$y_t = \mu_t + e_t, \qquad e_t \sim N\left(0, \sigma_e^2\right), \tag{1}$$

$$\mu_{t+1} = \mu_t + \eta_t, \qquad \eta_t \sim N\left(0, \sigma_\eta^2\right), \tag{2}$$

Where $\{e_t\}$ and $\{\eta_t\}$ are 2 independent Gaussian white noise series and $t = 1, \ldots, T$

The initial value μ_1 is either *given or follows a known distribution* and is independent of $\{e_t\}$ and $\{\eta_t\}$ for t>0.

Here μ_t is is a pure *random walk* with initial value μ_1 and y_t is an observed version of μ_t with added noise e_t μ_t is referred to as the *trend* of the series which is not directly observable, and y_t is the observed data with observational noise e_t.

The models above can be used to analyse the *realised* volatility of an asset price, where μ_t represents the underlying log volatility of the asset price and y_t is the log of realised volatility.

The model is a special linear *gaussian state space model*, with the variable μ_t called the *state* of the system at time t (not directly observed).

The y model provides the link between the data y_t and the state μ_t and is called the *observation equation* with measurement error e_t

The next μ_{t+1} governs the time evolution of the state variable and is the state equation with innovate η_t.

If $\sigma_e = 0$ then $y_t = \mu_t$ and there is no measurement error, which is an ARMA (0, 1, 0) model

If $\sigma_e > 0$ then there exist measurement error and y_t is an ARMA(0, 1, 1) model satisfying

$$(1-B)y_t = (1-\theta B)a_t \tag{3}$$

where $\{a_t\}$ is a gaussian white noise with mean zero and variance σ_a^2

Then, the values of θ and σ_{eta} are determined by σ_e and σ_η

From the initial model we have

$$\begin{aligned} (1-B)\mu_{t+1} = \eta_t \quad \text{or} \quad \mu_{t+1} = \frac{1}{1-B}\eta_t \\ \text{then we can rewrite} \quad y_t = \mu_t + e_t = y_t = \frac{1}{1-B}\eta_{t-1} + e_t \end{aligned} \tag{4}$$

And multiplying by B we have

$$(1-B)y_t = \eta_{t-1} + e_t - e_{t-1}$$

Then letting $(1-B)y_t = w_t$ we have $w_t = \eta_{t-1} + e_t - e_{t-1}$ And under the model assumptions it is easy to see that w_t is gaussian, $\text{Var}(w_t) = 2\sigma_e^2 + \sigma_\eta^2$ and $\text{Cov}(w_t, w_{t-1}) = -\sigma_e^2$ and $\text{Cov}(w_t, w_{t-j}) = 0$ for j>1

Then consequently w_t follows an MA(1) model and can be written as $w_t = (1-\theta B)a_t$

And by equating the variance and lag-1 autocovariance of $w_t = (1-\theta B)a_t = \eta_{t-1} + e_t - e_{t-1}$ and we have

$$(1+\theta^2)\sigma_a^2 = 2\sigma_e^2 + \sigma_\eta^2$$
$$\theta\sigma_a^2 = \sigma_e^2$$

Then for a given σ_e^2 and σ_η^2 we consider the ratio of these to form a quadratic function of θ, having 2 solutions which we select the one that satisfies $|\theta|<1$.

Idea is that if you have state equation with large variance, you wont be able to recover much.

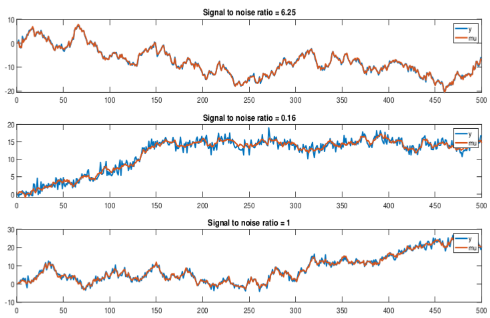

Figure 1: State Space Model

If signal to noise ratio = 0.16, observe blue try to recover red. It is not very informative, if ratio is 6 then signal is very informative

Kalman Filter

The aim of the analysis is to infer properties of the state μ_t alone from the data and the model. Let $F_t = \{y_1, \ldots, y_t\}$ be the information available at time t (inclusive) and assume that the model is known, including all parameters.

Three estimates of interest

1. Filtering : recover μ_t (remove measurement error)
2. Smoothing : estimate μ_t given all available information up to time T
3. Prediction : forecast μ_{t+k}

Analogy - filtering is figuring out the word you are reading based on knowledge accumulated from the beginning of the note, predicting is to guess the next word and smoothing is to decipher a particular word once you have read through the note.

Properties of Multivariate Normal Distribution Considering a multivariate normal distribution

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy} & \Sigma_{yy} \end{pmatrix}\right)$$

Kalman filter is tool which characterises the conditional distribution of μ_t given the data. Given the date we observe what is the distribution of μ_t

Notation

Let $\mu_{t|j} = E[\mu_t|F_j]$ and $\Sigma_{t|j} = \mathrm{Var}(\mu_t|F_j)$ be the conditional mean and variance of μ_t given F_j. $y_{t|j}$ is the conditional mean of y_t given F_j

And $v_t = y_t - y_{t|t-1}$ and $V_t = \mathrm{Var}(v_t|F_{t-1})$ be the 1 step ahead forecast error and its variance of y_t given F_{t-1}

The forecast error v_t is independent of F_{t-1} so that the conditional variance is the same as the unconditional variance, that is $\mathrm{Var}(v_t|F_{t-1}) = \mathrm{Var}(v_t)$

Then

$$Y_{t|t-1} = E[y_t|F_{t-1}] = E[\mu_t + e_t|F_{t-1}] = E[\mu_t|F_{t-1}] = \mu_{t|t-1}$$

And consequently,

$$v_t = y_t - y_{t|t-1} = y_t - \mu_{t|t-1}$$

and

$$V_t = \mathrm{Var}(y_t - \mu_{t|t-1}|F_{t-1}) = \mathrm{Var}(\mu_t + e_t - \mu_{t|t-1}|F_{t-1})$$

$$= \mathrm{Var}((\mu_t - \mu_{t|t-1}|F_{t-1})) + \mathrm{Var}(e_t|F_{t-1}) = \Sigma_{t|t-1} + \sigma_e^2$$

And then it is easy to see that

$$E[v_t] = E[E[y_t - y_{t|t-1}]|F_{t-1}] = E[y_{t|t-1} - y_{t|t-1}] = 0$$

$$\mathrm{Cov}(v_t, y_j) = E[v_t, y_j] = E[E[v_t y_j|F_{t-1}]] = E[y_j|E[v_t|F_{t-1}]] = 0, \qquad j < t$$

Then as expected the 1 step ahead forecast error is uncorrelated with $y_j$for j<t. And furthermore for the linear model in eq. (1) and eq. (2) $\mu_{t|t} = E[\mu_t|F_t] = E[\mu_t|F_{t-1}, v_t]$ and $\Sigma_{tZt} = \mathrm{Var}(\mu_t|F_t) = \mathrm{Var}(\mu_t|F_{t-1}, v_t)$

That is, the information set f_t can be written as $F_t = \{F_{t-1}, y_t\}$

Theorem 1 :
Properties of MV normal distribution useful to the Kalman filter under normality Suppose that x, y and z are 3 RV such that their joint distribution is MV normal, additionally assume that the diagonal block covariance Σ_{ww} is non singular for $w = x, y, z$ and $\Sigma_{yx} = 0$, then

1. $E[x|y] = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y)$
2. $\text{Var}(x|y) = \Sigma_{xx} - \Sigma_{xx}\Sigma_{yy}^{-1}\Sigma_{yx}$
3. $E[x|y,z] = E[x|y] + \Sigma_{xz}\Sigma_{zz}^{-1}(z - \mu_z)$
4. $\text{Var}(x|y,z) = \text{Var}(x|y) - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_i$

Then the conditional distribution of x given y is

$$x|y \sim\sim \mathcal{N}(\mu_x + \Sigma_x y \Sigma_y y^{-1}(y - \mu_y),\, \Sigma-)$$

$$\begin{pmatrix} \mu_t \\ \nu_t \end{pmatrix} |_{\mathcal{F}_{t-1}}$$

Goal is the conditional distribution $\mu_t|F_t$ based on new data y_t and the conditional distribution $\mu_t|F_{t-1}$

$$\begin{pmatrix} \mu_t \\ \nu_t \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_{t|t-1} \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{t|t-1} & \Sigma_{t|t-1} \\ \Sigma_{t|t-1} & V_t \end{pmatrix}\right)$$

Prediction

Initialise idea on conditional mean and variance of unobserved latent state variable (signals), take data minus initial value/expectation.

Look at forecast error variance

How does it work?

First remove measurement error then estimate μ_t given all available information, then forecast. Distribution of μ_t given information set $\mathcal{F}_t$ today. In order to today recover the value of μ_t need to update conditional expectation so that take into account signal to noise ratio. How much new noise contributes to the conditional variance expectation.

Filter latent process based on information $t - 1$ then we update forecast once new information has arrived. Kalman gain measures how much information does the new shock at time t add to uncertainty (?). Dent take information as given y_t has noise itself e so we only update conditional expectations proportionally to the signal to noise ratio.

Recover, then smoothing re estimating μ_t (trying to mitigate effect of starting values), then based on this forecast latent process. All based on one property of MVR norm.

After we know this we can write it down given this formula

Major idea of KF is to write down some expectations of latent process, then update these according to Kalman gain which measures model uncertainty plus new variance originating from noisy data. Which are inherently small in financial data. New data is not very informative (nothing in autocorrelation structure), so strongly depends on starting values. These values are not eaten up by new data as they may in physics.

[11]

Kalman Filter

The goal of the Kalman filter is to update knowledge of the state variable recursively when a new data point becomes available. That is, knowing the conditional distribution of μ_t given F_{t-1} and the new data y_t, we would like to obtain the conditional distribution of μ_t given F_t where as before $F_j = \{y_1 \dots, y_j\}$ since $F_t = \{F_{t-1}, v_t\}$ giving y_t and F_{t-1} is equivalent to giving v_t and F_{t-1}.

To derive the KF, it suffices to consider the joint conditional distribution of $(\mu_t, v_t)'$ given F_{t-1} before applying the above theorem

From the definition [**11.1.2**]

1 Tutorial 6

[PS6]

Is the signal to noise ratio the same no matter the number of simulations?

SNR is defined as $\frac{\sigma_e^2}{\sigma_\eta^2}$, the variances determine this, here they do not depend on anything, it is always defined by the variance of the error term in state equation and in the observation model.

B) mu and filter s

Our forecast error is defined as in [**slide 11**]

To understand the code, write the recursions using this slide

```
for (t in 1:T){
  predict_mu[t] = filter_mu[t]
  predict_S[t]= filter_S[t]
  v[t] = y[t]-predict_mu[t]
  V[t] = predict_S[t]+s_e^2
  K[t] = predict_S[t]/V[t]
  filter_mu[t+1] = predict_mu[t] + K[t]*v[t]
  filter_S[t+1] = predict_S[t]*(1-K[t])+s_eta^2
  print(V[t])
}
```

Old conditional expectation $\mu_{t|t+1} = \mu_{t|t+1} + K_t \cdot \nu_t$

In the for loop we take our conditional predictions and feed them 1 step ahead in the next iteration, so we have filtered out in iteration t becomes a prediction in $t+1$ so $\nu_{t+1} = y_{t+1} i \nu_{t|t+1}$

Kalman filter updating - the filtered at t becomes a prediction for t+1

Also possible to start at 2 and do filtering at t-1

What are the filter initialisations?

Can also draw from MV norm, this is local linear trend model with strong SNR, starting values matter less here, though this is the usual way to initialise.

Filter gets updated proportional to Kalman gain.

$$\begin{aligned} \nu = y_1 - u_{1|0} = y_1 - \ predictnu[1] \\ V_1 = predictS[1] + \sigma_e^2 \\ K_1 = \dots \\ filter[1] = predict.mu[1] + V[1] \cdot K[1] \\ \text{then for t = 2} \end{aligned}$$

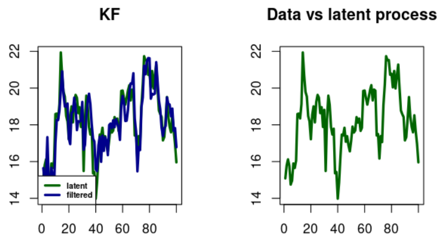

Figure 2

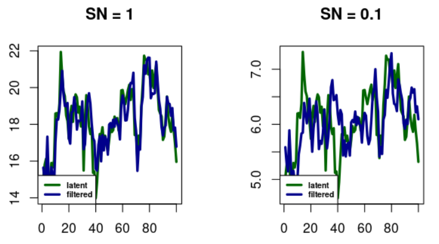

Figure 3

CI

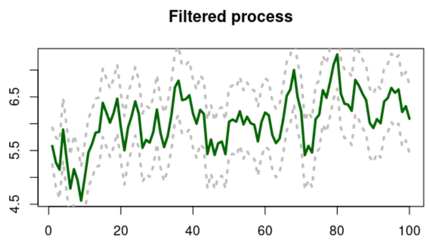

Figure 4

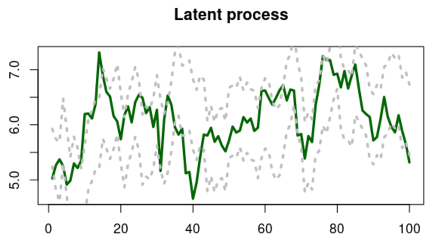

Figure 5

Fundamental property / assumption on μ_t, we assumed distribution given is **normal** and thus once we obtain filtration we can use this to write down CI. Conditional variance is given by filtered Σ.

Hypothetical Exam Question Why would CI be much wider where $\sigma_e^2 = 0.9$ and $\sigma_q^2 = 0.3$ than $\sigma_e^2 = 0.9$ and $\sigma_q^2 = 0.9$

The second has a SNR of 1, so higher ratio means the closer the observation and state equations are. Thinking about how CI are calculated, $\pm 2 \cdot \Sigma_{t|t}$ or $\Sigma_{t|t} = \Sigma_{t|t} + \sigma_q^2$

[**recursive slide**] - what about Kalman gain, need to think about how σ_η^2 influences variance, what does increase in σ_e^2, this is in denominator of Kalman gain :

This is all a recursive process

Why do we opt to calculate negative log likelihood?

```
 kf_loglik = function(y,s_e,s_eta){
 fit = kf_recursions(y,s_e,s_eta)
 # compute the negative log likelihood
 l = 0.5*log(2*pi)+0.5*(log(fit$V))+ 0.5*((fit$v^2)/fit$V)
 ll =sum(l)
 return(ll)
}
```

Estimates are RV, they can lie anywhere so have to do negative.