FPAS Uncertainties

Tomas Motl, OGResearch, April 2021

Introduction

Let's talk about initial condition.

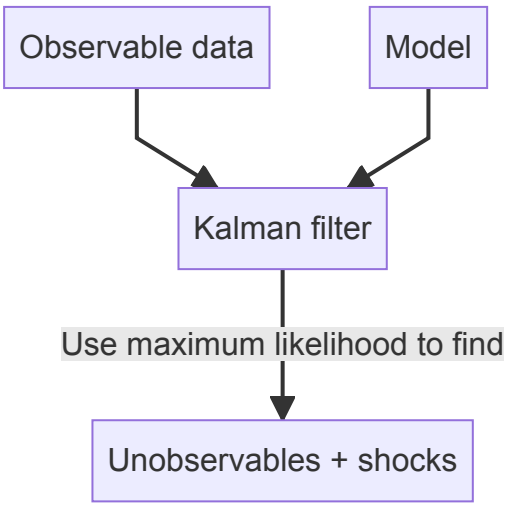
Recursive model solution:

$$X_t = A \cdot X_{t-1} + \epsilon_t$$

How do we choose X_0? Using only data is not enough, some variables are not observable ($\hat{y}_t, \hat{z}_t, \ldots$)

We use multivariate Kalman filter. Initial condition is always under our control. We need to help Kalman filter to get things right.

Kalman Filter



We need to expand the set of equations:

$$X_t = AX_{t-1} + \epsilon_t$$
$$Y_t = C_t X_t + \nu_t$$

Note that set of observable variables depends on particular periods, hence C_t, not C.

Check also paper "understanding_the_basis_of_the_kalman_filter.pdf" which is in on the Wiki.

Step 1: Predict

Prediction = forecast by the model.

Assume we know X_{t-1}. What's the best prediction for X_t?

$$X_{t|t-1} = AX_{t-1}$$
$$P_{t|t-1} = AP_{t-1|t-1}A^T + Q_t$$

$X_{t|t-1}$ is our estimate of X_t based on information available in time $t-1$.

$P_{t|t-1}$ is covariance matrix (uncertainty) of our prediction $X_{t|t-1}$.

Q_t is the covariance matrix of vector ϵ_t (**variance decomposition!**). Note that our uncertainty about X increases with prediction step.

KF gives us the whole joint probability distribution for all variables.

Step 2: Update

We obtain measurements Y_t and update our prediction $X_{t|t-1}$ to get our final estimate $X_{t|t}$.

How to combine our prediction and the measurement? Weight them together using the Kalman gain:

$$X_{t|t} = X_{t|t-1} + K_t(Y_t - CX_{t|t-1})$$
$$K_t = P_{t|t-1}C_t^T(C_t P_{t|t-1} C_t^T + R_t)^{-1}$$

where R_t is covariance matrix of vector η_t

Kalman filter weighs together prediction and information from measurements. The (inverse) weights are $P_{t|t-1}$ and R_t.

If we reduce matrices to scalars and we set uncertainty of measurement to zero ($R_t = 0$), we get

$$K_t = C_t^{-1}$$
$$X_{t|t} = X_{t|t-1} + C_t^{-1}(Y_t - CX_{t|t-1}) = X_{t|t-1} + C_t^{-1}Y_t - C_t^{-1}C_t X_{t|t-1} = C_t^{-1}Y_t$$

So we completely disregard predictions. If the measurements are completely precise, model prediction is not important.

Measurements reduce uncertainty about X_t, to the extent we trust them (K):

$$P_{t|t} = P_{t|t-1} - K_t C_t P_{t|t-1}$$

Shocks

Notice that

$$
\begin{aligned}
X_{t|t} &= C_t^{-1} Y_t \\
X_{t|t} &= A X_{t-1|t-1} + \epsilon_t \\
\epsilon_t &= C_t^{-1} Y_t - A X_{t-1|t-1}
\end{aligned}
$$

Shocks explain the prediction error, deviation between model prediction and observed data.

Shocks are important. They tell us:

1. How the model interprets data (which shocks explain data).
2. Where the model has problem interpreting data (large, autocorrelated shocks).

Application in macromodeling

In macroeconomic models, we usually don't introduce measurement errors and we set $R_t = 0$. So why do we need Kalman filter?

Unobservables?

Kalman filter is useful because not all variables are observed. Typically we have much more endogenous variables and shocks than measurements. Matrix C is not square, but rectangular.

Kalman filter will give us **the most likely** combination of shocks that explains data. If we know the most likely realization of shocks, we can also calculate unobservables.

Example

$$
\begin{aligned}
\hat{v}_t &= 0.5 \cdot \hat{v}_{t-1} + \varepsilon_t^1 \\
\overline{v}_t &= 0.5 \cdot \overline{v}_{t-1} + \varepsilon_t^2 \\
v_t &= \hat{v}_t + \overline{v}_t
\end{aligned}
$$

Here $X_t = [\overline{v}, \hat{v}, v]^T$ and $Y_t = [v_t^{obs}]$.

Let

$$
\begin{aligned}
v_0^{obs} &= 0 \\
v_1^{obs} &= 1
\end{aligned}
$$

There are infinitely many combinations of $\hat{v}_{t-1}, \overline{v}_{t-1}$ that fit the equations above. We need some decision criterion.

We specify assumptions about shock distribution:

$$
\begin{aligned}
\varepsilon^1 &\sim N(0, \sigma_1^2 = 2^2) \\
\varepsilon^2 &\sim N(0, \sigma_2^2 = 1^2)
\end{aligned}
$$

Maximum Likelihood, Kalman Smoother

This section is here to give you understanding how Kalman filter works. All of the operations below are embodied by the matrix algebra presented above.

The "smallest" combination of shocks is the most likely, so we minimize the expression:

$$\left(\frac{\varepsilon_1^1}{\sigma_1}\right)^2 + \left(\frac{\varepsilon_1^2}{\sigma_2}\right)^2$$

In real world applications, we optimize over all shocks and **all periods** (**Kalman smoother**). The optimization problem setup is:

$$\min\left[\left(\frac{\varepsilon_1^1}{\sigma_1}\right)^2 + \left(\frac{\varepsilon_1^2}{\sigma_2}\right)^2 + \ldots + \left(\frac{\varepsilon_T^1}{\sigma_1}\right)^2 + \left(\frac{\varepsilon_T^2}{\sigma_2}\right)^2\right]$$

subject to constraints (model + measurement):

$$X_t = AX_{t-1} + \epsilon_t$$
$$Y_t = CX_t + \nu_t$$

Minimizing the above = maximizing likelihood. Integral part of all your fancy estimation methods.

Kalman filter/smoother can be viewed as **least squares method** that finds the minimal shocks needed to explain data.

Numerical example

Recall our model

$$\hat{v}_t = 0.5 \cdot \hat{v}_{t-1} + \varepsilon_t^1$$
$$\overline{v}_t = 0.5 \cdot \overline{v}_{t-1} + \varepsilon_t^2$$
$$v_t = \hat{v}_t + \overline{v}_t$$

Here $X_t = [\hat{v}, \overline{v}, v]^T$ and $Y_t = [v_t^{obs}]$.

Transition matrix

$$A = \begin{pmatrix} 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0.5 & 0.5 & 0 \end{pmatrix}$$

Matrix of measurements:

$$C = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$$

Vector of shocks

$$\epsilon_t = \begin{pmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \\ 0 \end{pmatrix}$$

Uncertainty of initial condition:

$$P_0 = \begin{pmatrix} 2^2 & 0 & 0 \\ 0 & 1^2 & 0 \\ 0 & 0 & 1^2 + 2^2 \end{pmatrix}$$

Covariance matrix of shock vector ϵ:

$$Q = \begin{pmatrix} 2^2 & 0 & 0 \\ 0 & 1^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Measurements are precise:

$$R = 0$$

Let

$$v_0^{obs} = 0$$
$$v_1^{obs} = 1$$

Initialize variables at their means:

$$\hat{v}_0 = 0$$
$$\overline{v}_0 = 0$$
$$v_0 = 0$$

1) Prediction step:

$$X_{1|0} = AX_0 = A \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Uncertainty:

$$P_{1|0} = AP_0A^T + Q = \begin{pmatrix} 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0.5 & 0.5 & 0 \end{pmatrix} \begin{pmatrix} 2^2 & 0 & 0 \\ 0 & 1^2 & 0 \\ 0 & 0 & 1^2 + 2^2 \end{pmatrix} \begin{pmatrix} 0.5 & 0 & 0.5 \\ 0 & 0.5 & 0.5 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 2^2 & 0 & 0 \\ 0 & 1^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$P_{1|0} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0.25 & 0.25 \\ 1 & 0.25 & 1.25 \end{pmatrix} + \begin{pmatrix} 2^2 & 0 & 0 \\ 0 & 1^2 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 5 & 0 & 1 \\ 0 & 1.25 & 0.25 \\ 1 & 0.25 & 1.25 \end{pmatrix}$$

2) Update step

$$K = P_{1|0}C^T[CP_{1|0}C^T + R]^{-1}$$

$$K = \begin{pmatrix} 5 & 0 & 1 \\ 0 & 1.25 & 0.25 \\ 1 & 0.25 & 1.25 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \left[\begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 5 & 0 & 1 \\ 0 & 1.25 & 0.25 \\ 1 & 0.25 & 1.25 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} + 0 \right]^{-1}$$

$$K = \begin{pmatrix} 5 & 0 & 1 \\ 0 & 1.25 & 0.25 \\ 1 & 0.25 & 1.25 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} [5]^{-1} = \begin{pmatrix} 0.8 \\ 0.2 \\ 1 \end{pmatrix}$$

Recall that

$$X_{t|t} = X_{t|t-1} + K_t(Y_t - CX_{t|t-1})$$

$$X_{1,1} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0.8 \\ 0.2 \\ 1 \end{pmatrix} \left(1 - \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right) = \begin{pmatrix} 0.8 \\ 0.2 \\ 1 \end{pmatrix}$$

We received estimate of all three variables, despite observing only one.

Implied values of shocks:

$$\epsilon_1 = \begin{pmatrix} 0.8 \\ 0.2 \\ 0 \end{pmatrix}$$

Note that the first shock is larger than the second one, exactly in line with our assumptions about their standard errors.

Recall that

$$P_{t|t} = P_{t|t-1} - K_t C_t P_{t|t-1}$$

Numerically:

$$P_{t|t} = \begin{pmatrix} 5 & 0 & 1 \\ 0 & 1.25 & 0.25 \\ 1 & 0.25 & 1.25 \end{pmatrix} \begin{pmatrix} 0.8 \\ 0.2 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 5 & 0 & 1 \\ 0 & 1.25 & 0.25 \\ 1 & 0.25 & 1.25 \end{pmatrix} = \begin{pmatrix} 4.2 & -0.2 & 0 \\ -0.2 & 1.2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Note:

- uncertainty about first two variables ($\overline{v}$, $\hat{v}$) increased
- uncertainty about the third variable v remains zero (we measured it precisely)