4 Filtering

Filtering, which might be termed 'backcasting' is adjusting our beliefs about the state vector at $t-k$ using all of the available data, i.e. D_t. We would expect D_t to convey more information about θ_{t-k} than just D_{t-k}, as $D_{t-k} \subset D_t$. Each time we add a new datum, therefore, we can (if we so choose) revise our beliefs about previous values of the state vector.

A useful side-calculation. For reasons that will become apparent, we start by considering the mean and variance of $\theta_{t-k} \mid \theta_{t-k+1}, D_{t-k}$ for some $0 < k \leq t$. We can use the Bayes linear approach to compute this, once we have computed the mean and covariance structure of $\theta_{t-k}, \theta_{t-k+1} \mid D_{t-k}$. (Remember that the Bayes linear adjustment 'jumps' the θ_{t-k+1} from the left to the right of the bar.)

We already know that $\theta_{t-k} \mid D_{t-k} \sim \langle m_{t-k}, C_{t-k} \rangle$, where we computed m_{t-k} and C_{t-k} 'on the way up' (i.e. in the process of adjusting θ_t by the data D_t). Similarly, we have also computed the quantities

$$\mathsf{E}[\theta_{t-k+1} \mid D_{t-k}] =: a_{t-k+1} \quad \text{and} \quad \mathsf{Var}\,[\theta_{t-k+1} \mid D_{t-k}] =: R_{t-k+1}$$

'on the way up'. The only not-yet-computed term is

$$\mathsf{Cov}[\theta_{t-k}, \theta_{t-k+1} \mid D_{t-k}] = \mathsf{Cov}[\theta_{t-k}, \mathsf{E}[\theta_{t-k+1} \mid \theta_{t-k}] \mid D_{t-k}] = C_{t-k}\, G_{t-k+1}^{\mathsf{T}}.$$

Using the Bayes linear adjustment formulae we find that

$$\begin{aligned} \mathsf{E}[\theta_{t-k} \mid \theta_{t-k+1}, D_{t-k}] &= m_{t-k} + B_{t-k}\,(\theta_{t-k+1} - a_{t-k+1}) \\ \mathsf{Var}\,[\theta_{t-k} \mid \theta_{t-k+1}, D_{t-k}] &= C_{t-k} - B_{t-k}\, R_{t-k+1}\, B_{t-k}^{\mathsf{T}}, \end{aligned}$$

where $B_{t-k} := C_{t-k}\, G_{t-k+1}^{\mathsf{T}}\, R_{t-k+1}^{-1}$. Note how we have re-written the adjusted variance in a form that will be convenient below. These two quantities are crucial in what follows.

Computing the filtered mean and variance. Now we consider the mean and variance of $\theta_{t-k} \mid D_t$. We will write, in general, $\theta_{t-k} \mid D_t \sim \langle a_t(-k), R_t(-k) \rangle$, which is consistent with our forecasting notation. As initial conditions we must have $a_t(0) = m_t$ and $R_t(0) = C_t$.

Suppose (adopting the inductive approach) we have computed the two quantities $a_t(-k+1)$ and $R_t(-k+1)$. If we look at the graph of the DLM we see that θ_{t-k+1} separates θ_{t-k} from $y_{t-k+1}, \ldots, y_t$. In other words we can proceed by introducing θ_{t-k+1} and then dropping the later data:

$$\begin{aligned} \mathsf{E}[\theta_{t-k} \mid D_t] &= \mathsf{E}[\mathsf{E}[\theta_{t-k} \mid \theta_{t-k+1}, D_t] \mid D_t] && \text{introducing } \theta_{t-k+1} \\ &= \mathsf{E}[\mathsf{E}[\theta_{t-k} \mid \theta_{t-k+1}, D_{t-k}] \mid D_t] && \text{there go } y_{t-k+1}, \ldots, y_t \\ &= \mathsf{E}\big[m_{t-k} + B_{t-k}\,\{\theta_{t-k+1} - a_{t-k+1}\} \mid D_t\big] && \text{from above} \\ &= m_{t-k} + B_{t-k}\,\big\{\mathsf{E}[\theta_{t-k+1} \mid D_t] - a_{t-k+1}\big\} && \theta_{t-k+1} \text{ only unknown} \\ &= m_{t-k} + B_{t-k}\,\big\{a_t(-k+1) - a_{t-k+1}\big\}. \end{aligned}$$

For the variance, we follow the same steps:

$$\begin{aligned} \mathsf{Var}\,[\theta_{t-k} \mid D_t] = {} & \mathsf{Var}\,[\mathsf{E}[\theta_{t-k} \mid \theta_{t-k+1}, D_{t-k}] \mid D_t] + \\ & \quad \mathsf{E}[\mathsf{Var}\,[\theta_{t-k} \mid \theta_{t-k+1}, D_{t-k}] \mid D_t] \\ = {} & \mathsf{Var}\,[m_{t-k} + B_{t-k}\,(\theta_{t-k+1} - a_{t-k+1}) \mid D_t] + \\ & \quad \mathsf{E}\big[C_{t-k} - B_{t-k}\, R_{t-k+1}\, B_{t-k}^{\mathsf{T}} \mid D_t\big] \\ = {} & B_{t-k}\, \mathsf{Var}\,[\theta_{t-k+1} \mid D_t]\, B_{t-k}^{\mathsf{T}} + C_{t-k} - B_{t-k}\, R_{t-k+1}\, B_{t-k}^{\mathsf{T}} \\ = {} & C_{t-k} + B_{t-k}\,\big\{R_t(-k+1) - R_{t-k+1}\big\}\, B_{t-k}^{\mathsf{T}}. \end{aligned}$$

We summarise these two expressions for convenience. We have shown that $\theta_{t-k} \mid D_t \sim \langle a_t(-k), R_t(-k) \rangle$ for $0 < k \leq t$, where

$$\begin{aligned} a_t(-k) &= m_{t-k} + B_{t-k}\,\big\{a_t(-k+1) - a_{t-k+1}\big\} \\ R_t(-k) &= C_{t-k} + B_{t-k}\,\big\{R_t(-k+1) - R_{t-k+1}\big\}\, B_{t-k}^{\mathsf{T}} \end{aligned}$$

and $B_{t-k} := C_{t-k}\, G_{t-k+1}^{\mathsf{T}}\, R_{t-k+1}^{-1}$. You can see that the expressions for

$a_t(-k)$ and $R_t(-k)$ have a similar and rather pleasing structure: we start with our beliefs given D_{t-k}, and we modify these by some linear combination of what we thought about θ_{t-k+1} both before and after introducing the extra data $y_{t-k+1}, \dots, y_t$.

Computing the filtered covariance. We have not finished yet. There are still the covariances to be calculated: $C_t(-k, -j) := \mathsf{Cov}[\theta_{t-k}, \theta_{t-j} \mid D_t]$, for $0 \leq j < k \leq t$. We will show that

$$C_t(-k, -j) = B_{t-k}\, C_t(-k+1, -j)$$

with initial condition $C_t(-j, -j) = R_t(-j)$. Remembering that $t - k < t - j$, we introduce θ_{t-j} into the covariance to get

$$\mathsf{Cov}[\theta_{t-k}, \theta_{t-j} \mid D_t] = \mathsf{Cov}[\mathsf{E}[\theta_{t-k} \mid \theta_{t-j}, D_{t-j-1}], \theta_{t-j} \mid D_t]. \tag{1}$$

Now we must compute $\mathsf{E}[\theta_{t-k} \mid \theta_{t-j}, D_{t-j-1}]$. By introducing θ_{t-k+1} we find

$$\begin{aligned}
\mathsf{E}[\theta_{t-k} \mid \theta_{t-j}, D_{t-j-1}] &= \mathsf{E}[\mathsf{E}[\theta_{t-k} \mid \theta_{t-k+1}, D_{t-k}] \mid \theta_{t-j}, D_{t-j-1}] \\
&= \mathsf{E}[m_{t-k} + B_{t-k}\,(\theta_{t-k+1} - a_{t-k+1}) \mid \theta_{t-j}, D_{t-j-1}] \\
&= m_{t-k} + B_{t-k}\,(\mathsf{E}[\theta_{t-k+1} \mid \theta_{t-j}, D_{t-j-1}] - a_{t-k+1}).
\end{aligned}$$

This has taken us one step toward $t-j$. Applying the recursion repeatedly we will end up with an expression that looks like

$$\mathsf{E}[\theta_{t-k} \mid \theta_{t-j}, D_{t-j-1}] = \cdots + B_{t-k}\, B_{t-k+1} \cdots B_{t-j-1}\, \theta_{t-j} + \cdots$$

where only the bit in θ_{t-j} is relevant, because we are about to take the covariance and all the other bits are constant. Putting this back into (1), the result follows immediately.