2 Basic features, and forecasting

The basic structure. In the last lecture we set up a model with the following generic structure:

$$\begin{aligned} y_t \mid \theta_t &\sim \langle F_t^{\mathsf{T}} \theta_t,\, V_t \rangle \\ \theta_t \mid \theta_{t-1} &\sim \langle G_t\, \theta_{t-1},\, W_t \rangle \end{aligned} \tag{1}$$

where y_t is a scalar (the datum) and θ_t is a n-vector (the 'state vector'). The model is defined by the four quantities $\{F_t, V_t, G_t, W_t\}$ which we assume are known at time t: F_t is a n vector, V_t a non-negative scalar, G_t a $n \times n$ matrix, and W_t a $n \times n$ variance matrix, i.e. symmetric and positive (semi)-definite. Here '$\sim$' denotes 'is modelled as', and '$\langle \mu,\, \Sigma \rangle$' a random quantity with mean μ and variance Σ.

A more general way of writing this model is

$$\begin{aligned} y_t &= F_t^{\mathsf{T}} \theta_t + \nu_t & \nu_t &\sim \langle 0,\, V_t \rangle \\ \theta_t &= G_t\, \theta_{t-1} + \omega_t & \omega_t &\sim \langle \mathbf{0},\, W_t \rangle\,. \end{aligned} \tag{2}$$

Clearly, (2) implies (1). On the other hand, (2) is not a complete specification, as it does not tell us of the relationship between the quantities ν_t and ω_t, nor the relationship between, say, ν_t and ω_{t-1}.

The best way to think of a DLM is as the graphical structure

$$\begin{tikzcd} & y_1 \arrow[d, dash] & y_2 \arrow[d, dash] & & y_{t-1} \arrow[d, dash] & y_t \arrow[d, dash] & y_{t+1} \arrow[d, dash] & \\ \theta_0 \arrow[r, dash] & \theta_1 \arrow[r, dash] & \theta_2 \arrow[r, dotted, dash] & \text{etc.} \arrow[r, dotted, dash] & \theta_{t-1} \arrow[r, dash] & \theta_t \arrow[r, dash] & \theta_{t+1} \arrow[r, dotted, dash] & {} \end{tikzcd} \tag{3}$$

imposed on the model (2). This graph has two features. First, the state vector θ follows a markov process, so that $\theta_k \perp\!\!\!\perp \theta_j \mid \theta_{k-1}$, for all $1 \le j < k \le t$. This implies that $\omega_j \perp\!\!\!\perp \omega_k$ in (2). Second, θ_k separates y_k from everything else, so that $y_k \perp\!\!\!\perp y_j, \theta_j \mid \theta_k$. This implies that $\nu_k \perp\!\!\!\perp \omega_k$ and $\nu_k \perp\!\!\!\perp \nu_j, \omega_j$ in (2).

Notation. Our data are denoted by $D_k := (y_1, \dots, y_k)$, for $k = 1, 2, \dots, t$. The value of each y_k may be numeric, or it may indicate a missing value. Our prior beliefs about the state vector are written

$$\mathsf{E}[\theta_0] = m_0,\ \mathsf{Var}[\theta_0] = C_0 \quad \text{or, equivalently,} \quad \theta_0 \sim \langle m_0,\, C_0 \rangle\,.$$

We supply values for m_0 and C_0. In order to preserve the structure of the graph we must have $\theta_0 \perp\!\!\!\perp \nu_k, \omega_k$ for all $k = 1, \dots, t$. After adjusting by the data D_k we write

$$m_k := \mathsf{E}[\theta_k \mid D_k] \quad \text{and} \quad C_k := \mathsf{Var}[\theta_k \mid D_k]$$

for $k = 1, \dots, t$. The DLM allows us to compute m_k and C_k from m_{k-1}, C_{k-1} and y_k. In this way we can find, by advancing one period at a time, the quantities m_t and C_t, which are our current mean and variance for the state vector ('current' in the sense that we are using all currently available information). We will see how to do this later on.

Computing on the graph. The key operation on the graph is to simplify calculations of means and variances by introducing variables suggested by the graph. The fundamental result should be well-known to you.

Theorem 2.1. *If x, z and D are random quantities then the expectation of $x \mid D$ can be computed by introducing z, as $\mathsf{E}[x \mid D] = \mathsf{E}\big[\mathsf{E}[x \mid z, D] \mid D\big]$.*

We can use this result to forecast the mean of the state vector at a time $t+1$ starting from $\theta_t \mid D_t \sim \langle m_t,\, C_t \rangle$:

$$\begin{aligned} \mathsf{E}[\theta_{t+1} \mid D_t] &= \mathsf{E}[\mathsf{E}[\theta_{t+1} \mid \theta_t, D_t] \mid D_t] && \text{Introduce } \theta_t \\ &= \mathsf{E}[\mathsf{E}[\theta_{t+1} \mid \theta_t] \mid D_t] && \text{As } \theta_{t+1} \perp\!\!\!\perp D_t \mid \theta_t \text{ from the graph (3)} \\ &= \mathsf{E}[G_{t+1}\, \theta_t \mid D_t] && \text{From the model (2)} \\ &= G_{t+1}\, \mathsf{E}[\theta_t \mid D_t] && \text{Assuming } G_{t+1} \text{ known at time } t \\ &= G_{t+1}\, m_t && \text{By definition.} \end{aligned}$$

In forecasting we need to assume that $\{F_{t+k}, V_{t+k}, G_{t+k}, W_{t+k}\}$ are known at time t for $k = 1, 2, \dots$.

We can use Theorem 2.1 to extend this approach to the covariance.

Corollary 2.2. *Under the same conditions as Theorem 2.1,*

$$\mathsf{Cov}[x,\, y \mid D] = \mathsf{Cov}\big[\mathsf{E}[x \mid z, D],\, \mathsf{E}[y \mid z, D] \mid D\big] + \mathsf{E}\big[\mathsf{Cov}[x,\, y \mid z, D] \mid D\big]\,.$$

Proof. For simplicity take x, y and z to be scalars, and drop the conditioning on

D (which appears everywhere). Then

$$\begin{aligned}\mathsf{Cov}[x,\, y] &= \mathsf{E}[xy] - \mathsf{E}[x]\cdot\mathsf{E}[y] \\ &= \mathsf{E}[\mathsf{E}[xy \mid z]] - \mathsf{E}[\mathsf{E}[x \mid z]]\cdot\mathsf{E}[\mathsf{E}[y \mid z]]\end{aligned}$$

using Theorem 2.1. Subtracting and adding $\mathsf{E}\big[\mathsf{E}[x \mid z]\cdot\mathsf{E}[y \mid z]\big]$ allows this to be written in the required form. □

Remembering that the the variance is just a special case of the covariance, we can use this Corollary to compute the variance of $\theta_{t+1} \mid D_t$ from the starting point $\theta_t \mid D_t \sim \langle m_t,\, C_t\rangle$:

$$\begin{aligned}\mathsf{Var}\,[\theta_{t+1} \mid D_t] &= \mathsf{Var}\,[\mathsf{E}[\theta_{t+1} \mid \theta_t] \mid D_t] + \mathsf{E}[\mathsf{Var}\,[\theta_{t+1} \mid \theta_t] \mid D_t] \\ &= \mathsf{Var}\,[G_{t+1}\,\theta_t \mid D_t] + \mathsf{E}[W_{t+1} \mid D_t] \\ &= G_{t+1}\,C_t\,G_{t+1}^{\mathsf{T}} + W_{t+1}\end{aligned}$$

using the same steps as before, where G_{t+1}^{T} is the transpose of G_{t+1}. There is a very useful special case of Corollary 2.2, namely $\mathsf{Cov}[x,\, y \mid D] = \mathsf{Cov}[x,\, \mathsf{E}[y \mid x, D] \mid D]$. We can use this to compute $\mathsf{Cov}[\theta_t,\, \theta_{t+1} \mid D_t]$:

$$\begin{aligned}\mathsf{Cov}[\theta_t,\, \theta_{t+1} \mid D_t] &= \mathsf{Cov}[\theta_t,\, \mathsf{E}[\theta_{t+1} \mid \theta_t] \mid D_t] \\ &= \mathsf{Cov}[\theta_t,\, G_{t+1}\,\theta_t \mid D_t] \\ &= C_t\,G_{t+1}^{\mathsf{T}}.\end{aligned}$$

Note that in both of these results we have used implicitly the fact that $\theta_{t+1} \perp\!\!\!\perp D_t \mid \theta_t$, from (3).

General forecasting results. We can extend the above results to find, for any $k \geq 1$, that $\theta_{t+k} \mid D_t \sim \langle a_t(k),\, R_t(k)\rangle$ where

$$a_t(k) := G_{t+k}\,a_t(k-1) \quad \text{and} \quad R_t(k) := G_{t+k}\,R_t(k-1)\,G_{t+k}^{\mathsf{T}} + W_{t+k} \qquad (4)$$

subject to the initial values $a_t(0) := m_t$ and $R_t(0) := C_t$. We can also find, for any $0 \leq j < k$, that $\mathsf{Cov}[\theta_{t+j},\, \theta_{t+k} \mid D_t] = C_t(j,k)$, where

$$C_t(j,k) := C_t(j,k-1)\,G_{t+k}^{\mathsf{T}} \qquad (5)$$

subject to the initial values $C_t(j,j) = R_t(j)$.

It is also easy for us to forecast any linear combination of θ_{t+k}. For example, we are often interested in the 'mean response function' $\mu_{t+k} := F_{t+k}^{\mathsf{T}}\theta_{t+k}$. It follows immediately that

$$\mu_{t+k} \mid D_t \sim \big\langle F_{t+k}^{\mathsf{T}}\,a_t(k),\, F_{t+k}^{\mathsf{T}}\,R_t(k)\,F_{t+k}\big\rangle\,,$$

and, for any $0 < j \leq k$, that

$$\mathsf{Cov}[\mu_{t+j},\, \mu_{t+k} \mid D_t] = F_{t+j}^{\mathsf{T}}\,C_t(j,k)\,F_{t+k}.$$

Exercises

1. Prove Theorem 2.1 using elementary probability (for simplicity, do not bother to condition everything on D).
2. Prove $\mathsf{Cov}[x,\, y \mid D] = \mathsf{Cov}[x,\, \mathsf{E}[y \mid x, D] \mid D]$.
3. Prove (4) and (5).
4. Give explicit (i.e. non-recursive) expressions for (4) and (5) in the case where $\big\{F_{t+k}, V_{t+k}, G_{t+k}, W_{t+k}\big\} = \big\{F, V, G, W\big\}$ for all $k \geq 1$. A DLM in which these four quantities are time-invariant is known as a 'times series DLM'.