5 Simple time-series models

In this lecture we consider what might be called 'portmanteau' DLMs: that is, DLMs that we carry around because they are often found to be useful. Sometimes we have a detailed understanding of the process that underlies the data, and we embody this understanding in a careful choice of G_t in the state equation. Complex models of physical processes are often of this type. But at other times we want a simple flexible model structure that can accommodate a wide range of different 'types' of data, where we have less of a feel for how the data evolves through time. These models are time-invariant, i.e. we can write the quantities F, V, G and W without time subscripts.

The simplest way to classify portmanteau models is by their mean response function, $\mu_{t+k} := F^\intercal \theta_{t+k}$. A very useful class of models has the property that $\mathsf{E}[\mu_{t+k} \mid D_t]$ is a polynomial of given degree in k. The coefficients of the polynomial are functions of the data D_t through the adjusted mean $m_t := \mathsf{E}[\theta_t \mid D_t]$. Therefore as we collect more data we are adjusting our beliefs about the coefficients of the polynomial. This type of model is termed the *Polynomial growth model* (PGM).

Polynomial growth models. This section describes the PGM (which you need to know as it is very useful) and explains why it has the properties described above (which is less important for you to know, but serves as a good example of how the simple structure of the DLM can give rise to interesting predictive models).

A PGM of order n has $F = E_n$ and $G = L_n$, where

$$(E_n)_i = 1_{i=1} \qquad (L_n)_{ij} = 1_{i \le j}$$

for $i, j = 1, \dots, n$, where 1_p is the indicator function of proposition p. So E_n picks out the first term in the state vector, and L_n is a triangular matrix of 0s below the diagonal and 1s on and above it. To prove that this specification has a polynomial structure in k, and to find out what it is, we need the following interesting result (to be proved as an exercise).

Theorem 5.1. *Define* $\mathsf{C}^i_j := \dfrac{i!}{j!(i-j)!}$ (i.e. "i choose j"), *then*

$$S^m_j := \sum_{i=j}^{m} \mathsf{C}^i_j \quad \Longrightarrow \quad S^m_j = \mathsf{C}^{m+1}_{j+1}$$

for $0 \le j \le m$.

Using this result we can investigate

$$\mathsf{E}[\mu_{t+k} \mid D_t] = (E_n)^\intercal \mathsf{E}[\theta_{t+k} \mid D_t] = (E_n)^\intercal (L_n)^k \, m_t.$$

The vector $(E_n)^\intercal (L_n)^k$ is just the first row of $(L_n)^k$, so this is the only part of $(L_n)^k$ we need to worry about when investigating the mean response function. From the structure of L_n, we note that if the first row of $(L_n)^k$ is $(v_1, \dots, v_n)$, then the first row of $(L_n)^{k+1} = (L_n)^k L_n$ is $(v_1, v_1 + v_2, \dots, v_1 + \dots + v_n)$. Now we can prove that

$$(E_n)^\intercal (L_n)^k = \left(\mathsf{C}^{k-1}_{k-1}, \mathsf{C}^{k}_{k-1}, \dots, \mathsf{C}^{k+n-2}_{k-1} \right). \tag{1}$$

This is certainly true for $k = 1$, remembering that $0! = 1$ by definition, giving a vector of all 1s as required. Now suppose it is true for arbitrary $k \ge 1$. Then

$$\begin{aligned}
(E_n)^\intercal (L_n)^{k+1} &= \left(\mathsf{C}^{k-1}_{k-1}, \mathsf{C}^{k-1}_{k-1} + \mathsf{C}^{k}_{k-1}, \dots, \mathsf{C}^{k-1}_{k-1} + \dots + \mathsf{C}^{k+n-2}_{k-1} \right) \\
&= \left(S^{k-1}_{k-1}, S^{k}_{k-1}, \dots, S^{k+n-2}_{k-1} \right) \\
&= \left(\mathsf{C}^{k}_{k}, \mathsf{C}^{k+1}_{k}, \dots, \mathsf{C}^{k+n-1}_{k} \right)
\end{aligned}$$

using Theorem 5.1. Substituting $k+1$ for k in (1) we get the same answer, so (1) is proved by induction.

So we have found that, in general,

$$\mathsf{E}[\mu_{t+k} \mid D_t] = \sum_{\ell=1}^{n} \mathsf{C}_{k-1}^{k+\ell-2}\, m_{t\ell}$$

where $\mathsf{E}[\theta_t \mid D_t] = m_t = (m_{t1}, \dots, m_{tn})$. Each coefficient has the structure

$$\mathsf{C}_{k-1}^{k+\ell-2} = \frac{(k+\ell-2)!}{(k-1)!(\ell-1)!} = \begin{cases} 1 & \ell = 1 \\ \dfrac{k(k+1)\cdots(k+\ell-2)}{(\ell-1)!} & \ell = 2, \dots, n \end{cases}$$

which is a polynomial of degree $\ell - 1$ in k. When we add them altogether to find $\mathsf{E}[\mu_{t+k} \mid D_t]$ we have a polynomial of degree $n-1$ in k. But you can see that the coefficients on each $k^{(\ell-1)}$ term are quite complicated functions of m_t.

As examples, consider first the case $n = 1$, known as the "random walk". In this case we have

$$\mathsf{E}[\mu_{t+k} \mid D_t] = m_{t1}$$

a polynomial of degree 0 in k. The case $n = 2$ is often known, for obvious reasons, as the "linear trend" model. For $n = 2$ we have

$$\begin{aligned}\mathsf{E}[\mu_{t+k} \mid D_t] &= \mathsf{C}_{k-1}^{k-1}\, m_{t1} + \mathsf{C}_{k-1}^{k}\, m_{t2} \\ &= m_{t1} + k\, m_{t2}.\end{aligned}$$

For $n = 3$ ("quadratic trend" model) we have

$$\begin{aligned}\mathsf{E}[\mu_{t+k} \mid D_t] &= \mathsf{C}_{k-1}^{k-1}\, m_{t1} + \mathsf{C}_{k-1}^{k}\, m_{t2} + \mathsf{C}_{k-1}^{k+1}\, m_{t3} \\ &= m_{t1} + k\, m_{t2} + \frac{(k+1)k}{2}\, m_{t3} \\ &= m_{t1} + (m_{t2} + m_{t3}/2)\, k + (m_{t3}/2)\, k^2.\end{aligned}$$

We could go on, but you would have to have fairly strong prior beliefs to want to predict future data using a quartic, or higher, polynomial in k.

The origin of the polynomial growth model. The motivation for the polynomial growth model becomes apparent when the state equation is re-written as

$$\theta_{tj} = \begin{cases} \theta_{t-1,j} + \theta_{t,j+1} + \delta\theta_{tj} & j = 1, \dots, n-1 \\ \theta_{t-1,n} + \delta\theta_{tn} & j = n \end{cases} \tag{2}$$

(careful with the time subscripts here!) where $\delta\theta_t \sim \langle 0, D \rangle$ and D is a diagonal variance matrix, i.e. the components of $\delta\theta_t$ are uncorrelated. The polynomial growth model can be recovered in this representation by back-substitution (e.g. start by writing $\theta_{t,j+1}$ as $\theta_{t-1,j+1} + \theta_{t,j+2} + \delta\theta_{t,j+1}$). When we do this we accumulate the errors for each j, and so the error on the state equation is $\omega_t = L_n\, \delta\theta_t$, and the error variance on the state equation, which we call W, is of the form $(L_n)D(L_n)^{\mathsf{T}}$.

In (2), $\theta_{t,j+1}$ is a bit like the first derivative of θ_{tj} with respect to time. The role of $\delta\theta_t$ is to perturb these 'derivatives' with random disturbances of given variance. The first component of the state vector, θ_{t1} has $n-1$ of these derivatives, which is why, heuristically, the expectation of the mean response function is a polynomial of degree $n-1$.

Physical applications. In this form we can see why the PGM might have physical applications. Consider a body moving in a 1-dimensional Euclidean space, not subject to any forces. This means that in the future it should move with the same velocity that it has now, and its expected displacement $\mathsf{E}[\mu_{t+k} \mid D_t]$ will be a linear function of k. This suggests using a linear trend PGM (i.e. $n = 2$) to model the displacement of this body relative to a fixed origin. In this case θ_{t1} denotes the location and

θ_{t2} the change in location per time-step, or 'velocity':

$$\begin{aligned}\theta_{t1} &= \theta_{t-1,1} + \theta_{t2} + \delta\theta_{t1} \\ &= \theta_{t-1,1} + \theta_{t-1,2} + (\delta\theta_{t1} + \delta\theta_{t2}) \\ \theta_{t2} &= \theta_{t-1,2} + \delta\theta_{t2}\end{aligned}$$

noting that $\omega_t = L_2\,\delta\theta_t$. In matrix notation,

$$\theta_t = L_2\,\theta_{t-1} + \omega_t \qquad \omega_t \sim \langle \mathbf{0},\, L_2 D (L_2)^{\mathsf{T}} \rangle$$

You can see that δ_{t1} is the deviation of location θ_{t1} from its linear trend, and δ_{t2} is the deviation of 'velocity' θ_{t2} from its previous value. If you thought that constant velocity was a good model you would set both $D_{11} = \mathsf{Var}\,[\delta\theta_{t1}]$ and $D_{22} = \mathsf{Var}\,[\delta\theta_{t2}]$ small. If you thought that the force was not constant, so that the velocity could change unpredictably from period to period, you might want to make D_{22} larger. You initial beliefs $\theta_0 \sim \langle m_0,\, C_0 \rangle$ describe your initial uncertainty about location and 'velocity', before you collect any data.

The same body falling in a constant gravitational field has a fixed acceleration, and for this purpose a quadratic trend PGM ($n = 3$) would be appropriate. Exactly the same types of considerations would apply when choosing the 3 diagonal elements of D.

Bolting DLMs together. Following on from the above application, you may wonder how it is possible to model a body moving in a 2- or 3-dimensional space. The answer is simple: the direct sum of two (or more) DLMs is also a DLM: we just 'bolt' them together. For simplicity consider the case of a body moving in 2-dimensional space, not subject to any forces. We would probably want to use a linear trend model for each dimension, giving us a state vector of length 4 in all (two locations and two velocities). Let's write θ for the state vector of the x-dimension and ψ for the state vector the y-dimension. Then our combined state equation is

$$\begin{pmatrix}\theta_t \\ \psi_t\end{pmatrix} = \begin{pmatrix} L_2 & \mathbf{0} \\ \mathbf{0} & L_2 \end{pmatrix} \begin{pmatrix}\theta_{t-1} \\ \psi_{t-1}\end{pmatrix} + \begin{pmatrix}\omega_{t\theta} \\ \omega_{t\psi}\end{pmatrix}.$$

The mean of the ω_t term is zero, and the variance is

$$\mathsf{Var}\begin{bmatrix}\omega_{t\theta} \\ \omega_{t\psi}\end{bmatrix} = \begin{pmatrix} L_2\,D_\theta (L_2)^{\mathsf{T}} & \mathbf{0} \\ \mathbf{0} & L_2\,D_\psi (L_2)^{\mathsf{T}} \end{pmatrix}$$

where D_θ and D_ψ are the diagonal variance matrices. D_θ, for example, is our uncertainty about progress in the x-direction, and $(D_\theta)_{22}$ is our uncertainty about the constancy of the 'velocity' in the x-direction.

How do we use this 'bolted-together' joint state equation? Usually we consider collecting information about some function of the location of the body, which is (θ_{t1}, ψ_{t1}), and use this information to update our beliefs about the body's location and future trajectory. In the simplest example imagine 'pinging' the body with radar, to identify only its distance from the given location of the radar station $(\bar{x}, \bar{y})$,

$$r_t = \sqrt{(\theta_{t1} - \bar{x})^2 + (\psi_{t1} - \bar{y})^2} + \nu_t$$

where ν_t describes the accuracy of the radar measurment. We see that r_t is a non-linear function of the state vector. We will see how to handle non-linear obsevation equations in the next lecture.

An interesting application along these lines can be found in "The application of state estimation to target tracking", C. B. Chang and J. A. Tabaczynski, *IEEE Transactions on Automatic Control*, 1984, v. 29, pp. 98–109.