Chapter 3

Classifying and building DLMs

In this chapter we will consider a special form of DLM for which the transition matrix F_k and the observation matrix H_k are constant in time, i.e., $F_k = F$ and $H_k = H$ for any $k \geq 0$. This sort of DLM can be referred to as a *time-series* DLM. In this case, denoting $\boldsymbol{\mu}_n = \mathbb{E}(\boldsymbol{y}_n \,|\, \boldsymbol{\theta}_n) = H\boldsymbol{\theta}_n$ the conditional expectation of $\boldsymbol{y}_n$ given $\boldsymbol{\theta}_n$, the expected value of $\boldsymbol{\mu}_{k+\delta}$ for some $\delta \geq 0$ given observations up to time k can be expressed as a function of δ as

$$g_k(\delta) = \mathbb{E}(\boldsymbol{\mu}_{k+\delta} \,|\, \boldsymbol{y}_{0:k} = y_{0:k}) = HF^{\delta}\hat{m}_k,$$

where $\hat{m}_k$ is the updated mean at time k. The vectors $g_k(\delta)$, $\delta \geq 0$, can be referred to as *forecasts* and the function g_k is called the *forecast function.* Note that for any $d \times d$ matrix M, the convention is $M^0 = I_d$ with I_d the $d \times d$ identity matrix. We also assume that the dimension d' of the observation space $\mathcal{Y}$ is equal to 1, this type of DLM is called an *univariate* DLM; in this case, $\boldsymbol{y}_k$ and $\boldsymbol{\mu}_k$ are real-valued random variables and g_k is simply a real-valued function.

3.1 Classifying DLMs

Considering the fact that the class of all DLMs is extremely large, it is useful to restrict our attention to specific models that have attractive properties and to limit any redundancy between models. This is the objective in the following sections.

3.1.1 Observability

For the sake of simplicity, we first consider the case where there is no transition noise, that is where $\boldsymbol{u}_k = 0$ almost surely for any $k \geq 0$. It follows that $\boldsymbol{\theta}_k = F\boldsymbol{\theta}_{k-1}$ and $\boldsymbol{\mu}_n = HF^{n-k}\boldsymbol{\theta}_k$ for any $n \geq k$. Note in particular that any $\boldsymbol{\theta}_n$ is known as soon as $\boldsymbol{\theta}_k$ is given for some $k \leq n$ and the problem can be reduced to estimating $\boldsymbol{\theta}_0$. A natural question is: how many observations need to be collected for $\boldsymbol{\theta}_k$ to be completely determined? Since $\boldsymbol{\theta}_k$ is a vector of dimension d, at least d observations are required. We therefore consider $n = k + d - 1$ and denote by $\boldsymbol{\mu}_{k:n}^{\mathsf{T}}$ the (column) vector of the d successive conditional expectations of the observations, which can be related to $\boldsymbol{\theta}_k$ as follows:

$$\boldsymbol{\mu}_{k:n}^{\mathsf{T}} = T\boldsymbol{\theta}_k$$

with the $d \times d$ *observability matrix* T defined as

$$T = \begin{pmatrix} H \\ HF \\ \vdots \\ HF^{d-1} \end{pmatrix}.$$

For $\boldsymbol{\theta}_k$ to be fully determined by this equation, T has to be invertible (which is equivalent to T having full rank). If this is the case then $\boldsymbol{\theta}_k = T^{-1}\boldsymbol{\mu}_{k:n}^{\mathsf{T}}$.

These ideas can be extended to the general case where there is no specific assumption on the transition noise $\boldsymbol{u}_k$, which leads to the following definition.

Definition 3.1 (Observability)**.** A univariate time-series DLM is *observable* if and only if the associated observability matrix has full rank.

Unless otherwise stated, the DLMs considered in the remainder of this chapter will be assumed to be observable, and we denote by $\mathbb{M}$ the class of all observable univariate time-series DLMs. In some specific scenarios, it might be useful to consider DLMs that are unobservable, for instance if there exist equality constraints on the parameter vector $\boldsymbol{\theta}_k$ for any $k \geq 0$ in which case rank deficiency can be compensated.

3.1.2 Similar and equivalent models

The class $\mathbb{M}$ of all observable DLMs of interest is still very large and we might need to identify and remove any redundancy in it in order to further simplify it. In this section, we will introduce two ways of identifying models with related features, the first one will focus on the transition matrix whereas the second one will involve the entire DLM.

Throughout this section, we will consider two models $\mathcal{M}, \mathcal{M}' \in \mathbb{M}$ with respective transition matrices F and F', respective observation matrices H and H' and respective forecast functions g_k and g'_k.

Definition 3.2 (Similarity)**.** The models $\mathcal{M}$ and $\mathcal{M}'$ are *similar* if and only if F and F' have identical eigenvalues.

The main implication of the definition of similarity is that two similar models $\mathcal{M}$ and $\mathcal{M}'$ will have the same form of forecast functions. For instance, if F and F' both have d distinct eigenvalues $\lambda_1, \ldots, \lambda_d$, then there exist E and E' such that $F = E\Lambda E^{-1}$ and $F' = E'\Lambda E'^{-1}$ with Λ the diagonal matrix of eigenvalues. It follows that

$$g_k(\delta) = \sum_{i=1}^{d} a_{k,i}\lambda_i^{\delta} \qquad \text{and} \qquad g'_k(\delta) = \sum_{i=1}^{d} b_{k,i}\lambda_i^{\delta},$$

for some scalars $a_{k,i}$ and $b_{k,i}$, $i \in \{1, \ldots, d\}$, so that the forecast functions have the same algebraic form.

Another way of characterising similar models is via the existence of a non-singular $d \times d$ *similarity matrix* S verifying $F = SF'S^{-1}$. In this case, the matrices F and F' are also said to be similar. This way of relating transition matrices via a similarity matrix can be extended to the whole model, which will in turn allow for defining a more specific connection between DLMs.

Let the DLM $\mathcal{M}'$ be characterised by the equations

$$\boldsymbol{\theta}'_k = F'\boldsymbol{\theta}'_{k-1} + \boldsymbol{u}'_k \tag{3.1a}$$
$$\boldsymbol{y}_k = H'\boldsymbol{\theta}'_k + \boldsymbol{v}_k \tag{3.1b}$$

with $\boldsymbol{u}'_k \sim \mathrm{N}(\cdot\,; 0, U'_k)$ and $\boldsymbol{v}_k \sim \mathrm{N}(\cdot\,; 0, V_k)$. Given a non-singular $d \times d$ matrix S, we can re-parametrise the state $\boldsymbol{\theta}'_k$ of the DLM $\mathcal{M}'$ as $\boldsymbol{\theta}_k = S\boldsymbol{\theta}'_k$, so that the system of equations (3.1) can now be expressed as

$$\boldsymbol{\theta}_k = SF'S^{-1}\boldsymbol{\theta}_{k-1} + S\boldsymbol{u}'_k = SF'S^{-1}\boldsymbol{\theta}_{k-1} + \boldsymbol{u}_k$$
$$\boldsymbol{y}_k = H'S^{-1}\boldsymbol{\theta}_k + \boldsymbol{v}_k$$

with $\boldsymbol{u}_k \sim \mathrm{N}(\cdot\,; 0, SU'_kS^{\intercal})$. It follows that a new model $\mathcal{M}$ can be defined via the transition and observation matrices $F = SF'S^{-1}$ and $H = H'S^{-1}$ and via the noise covariances $U_k = SU'_kS^{\intercal}$ and V_k. Noting that

$$F^{\delta} = (SF'S^{-1})^{\delta} = (SF'S^{-1})(SF'S^{-1})\ldots(SF'S^{-1}) = SF'^{\delta}S^{-1},$$

for any $\delta \geq 0$, and that $\hat{m}_k = S\hat{m}'_k$, it is easy to verify that

$$g_k(\delta) = HF^{\delta}\hat{m}_k = (H'S^{-1})(SF'^{\delta}S^{-1})(S\hat{m}'_k) = H'F'^{\delta}\hat{m}'_k = g'_k(\delta),$$

that is, $\mathcal{M}$ and $\mathcal{M}'$ have the same forecast function. In particular, these models are obviously similar. Since models $\mathcal{M}$ and $\mathcal{M}'$ have been assumed to be observable, their observability matrices can be introduced as T and T' respectively, and we find that $T' = TS$. This gives an expression of the similarity matrix S from $\mathcal{M}'$ to $\mathcal{M}$ as $S = T^{-1}T'$. We can now define a more restrictive relation between models as follows.

Definition 3.3 (Equivalence)**.** Let $\mathcal{M}, \mathcal{M}' \in \mathbb{M}$ be two similar models with respective observability matrix T and T' and denote by $S = T^{-1}T'$ the associated similarity matrix. The models $\mathcal{M}$ and $\mathcal{M}'$ are said to be *equivalent* if $m_0 = Sm'_0$ and $P_0 = SP'_0S^\intercal$, and, for any $k \geq 0$,

$$U_k = SU'_kS^\intercal \qquad \text{and} \qquad V_k = V'_k.$$

This concept is illustrated in the following example.

Example 3.1. If we consider a tracking problem on the real line, then defining the parameter vector as $\boldsymbol{\theta}'_k = (\boldsymbol{x}_k, \dot{\boldsymbol{x}}_k)^\intercal$ with $\boldsymbol{x}_k$ the (random) position and $\dot{\boldsymbol{x}}_k$ the (random) velocity or as $\boldsymbol{\theta}_k = (\dot{\boldsymbol{x}}_k, \boldsymbol{x}_k)^\intercal$ obviously yields similar models. It is easy to see that

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

so that, for instance, the transition matrix F for the state $\boldsymbol{\theta}_k$ can be deduced from the transition matrix F' for $\boldsymbol{\theta}'_k$ as follows

$$F' = \begin{pmatrix} 1 & \Delta \\ 0 & 1 \end{pmatrix} \qquad \text{and} \qquad F = SF'S^{-1} = \begin{pmatrix} 1 & 0 \\ \Delta & 1 \end{pmatrix}$$

with Δ the duration of a time step, noting that $S^{-1} = S$ (matrices with this property are called *involutory*).

Now that we have defined some ways of identifying models with similar features, the next question is: how to pick the most natural model(s) out of an equivalence class of similar/equivalent models.

3.1.3 Canonical models

From the viewpoint of linear algebra, the relation of similarity between matrices is an equivalence relation on the space of square matrices. For a given equivalence class of similar matrices, it is then natural to consider the one with the "simplest" form as the *canonical* element in the class. For example, if a given equivalence class is made of diagonalisable matrices then it makes sense to consider the diagonal matrix it contains as canonical. However, not all matrices are diagonalisable so the best we can do is pick the "most" diagonal matrix in each equivalence class. This matrix is usually considered to be the one of the form

$$J_{n_{1:b}}(\lambda_1, \ldots, \lambda_b) = \begin{pmatrix} J_{n_1}(\lambda_1) & & \\ & \ddots & \\ & & J_{n_b}(\lambda_b) \end{pmatrix}$$

with b and $n_1, \ldots, n_b$ positive integers and with $\lambda_1, \ldots, \lambda_b$ scalars, where $J_n(\lambda)$ is a $n \times n$ matrix of the form

$$J_n(\lambda) = \begin{pmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix}$$

for any λ and any positive integer n. The matrix $J_{n_{1:b}}(\lambda_1, \ldots, \lambda_b)$ is said to be in the *Jordan form* and $J_n(\lambda)$ is called a *Jordan block*.

The ideas introduced here for square matrices can be extended to any DLMs in $\mathbb{M}$. Henceforth, we will denote by e_d the vector of dimension d defined by $e_d = (1, 0, \ldots, 0)$. We consider a DLM $\mathcal{M} \in \mathbb{M}$ with transition matrix F, observation matrix H, observability matrix T, noise matrices U_k and V_k and prior $\mathrm{N}(\cdot\,; m_0, P_0)$. As a first step, the following two special cases can be identified:

One real eigenvalue If the transition matrix F has a single real eigenvalue λ of multiplicity d then

- the matrix F is similar to $J_d(\lambda)$
- the first element of H is different from zero
- any model $\mathcal{M}'$ with transition matrix $J_d(\lambda)$ and observation matrix e_d is said to be **a** *canonical similar model*

Multiple real eigenvalues If the transition matrix F has multiple real eigenvalues $\lambda_1, \dots, \lambda_b$ of respective multiplicities $n_1, \dots, n_b$ then

- the matrix F is similar to $J_{n_{1:b}}(\lambda_1, \dots, \lambda_b)$
- let s be the sequence $(1, n_1+1, \dots, n_1+\dots+n_{b-1}+1)$, then the s_i-th element of H is different from zero for any $i \in \{1, \dots, b\}$
- any model $\mathcal{M}'$ with transition matrix $J_{n_{1:b}}(\lambda_1, \dots, \lambda_b)$ and observation matrix $(e_{n_1}, \dots, e_{n_b})$ is said to be **a** *canonical similar model*

In both cases, the canonical similar model $\mathcal{M}^*$ with noise matrices $SU_kS^\intercal$ and V_k, with $S = (T^*)^{-1}T$, and prior $\mathrm{N}(\cdot\,; Sm_0, SP_0S^\intercal)$ is said to be **the** *canonical equivalent model.* Illustrations of the behaviour of a DLM with one real eigenvalue are given in Figure 3.1.

Remark 3.1. When selecting a given observation matrix H^* for the canonical equivalent model, only the observability of the model needs to be ensured. Indeed, the similarity matrix S will ensure that observing the state $\boldsymbol{\theta}_k$ in model $\mathcal{M}$ via $H\boldsymbol{\theta}_k$ will be equivalent to observing the state $\boldsymbol{\theta}^*_k$ in model $\mathcal{M}^*$ via $H^*\boldsymbol{\theta}^*_k$. For instance, if $F = J_d(\lambda)$ for some given real λ and for $d = 2$ and if $H = (1, 1)$ then the canonical equivalent model also has $J_d(\lambda)$ as transition matrix (this is already the canonical form) and is such that $H^* = e_2$. To compute the corresponding similarity matrix, one must first compute the inverse of T^* as

$$(T^*)^{-1} = \begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ -\lambda & 1 \end{pmatrix}$$

so that

$$S = (T^*)^{-1}T = \begin{pmatrix} 1 & 0 \\ -\lambda & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ \lambda & \lambda+1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

The state $\boldsymbol{\theta}^*_k$ can then be expressed as a function of the components of $\boldsymbol{\theta}_k$ as

$$\boldsymbol{\theta}^*_k = S\boldsymbol{\theta}_k = \begin{pmatrix} \boldsymbol{\theta}_{k,1} + \boldsymbol{\theta}_{k,2} \\ \boldsymbol{\theta}_{k,2} \end{pmatrix},$$

which indeed ensures that observing $H\boldsymbol{\theta}_k$ is equivalent to observing $H^*\boldsymbol{\theta}^*_k$.

In general, not all eigenvalues will be real. However, since F is a real matrix, complex eigenvalues come in complex conjugate pairs. We consider the case $d = 2$ to illustrate the specific properties of such models. In this case, the two eigenvalues can be expressed as

$$\lambda_1 = \lambda \exp(i\omega) \qquad \text{and} \qquad \lambda_2 = \lambda \exp(-i\omega) \tag{3.3}$$

for some scalars λ and ω and with $i = \sqrt{-1}$. Although F is similar to the matrix

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix},$$

working with complex matrices is not generally appealing. Yet, a similarity matrix S can be identified to arrive to a real matrix $\lambda R(\omega)$ with

$$S = \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} \qquad \text{and} \qquad R(\omega) = \begin{pmatrix} \cos(\omega) & \sin(\omega) \\ -\sin(\omega) & \cos(\omega) \end{pmatrix}.$$

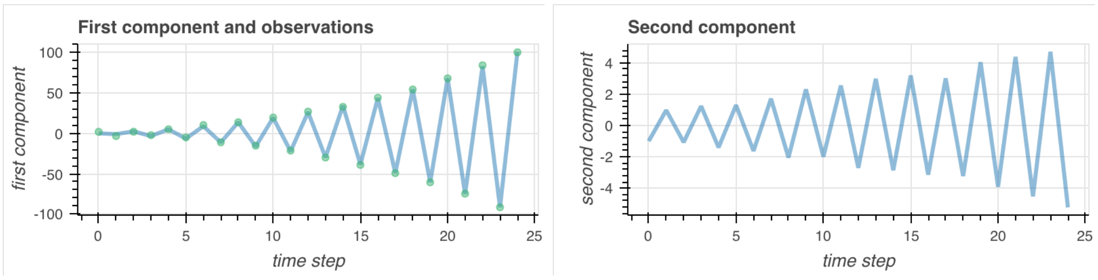
We can indeed verify that $S\Lambda S^{-1} = \lambda R(\omega)$; the corresponding observation matrix is

$$H = (1, 1)S^{-1} = (1, 0) = e_2.$$

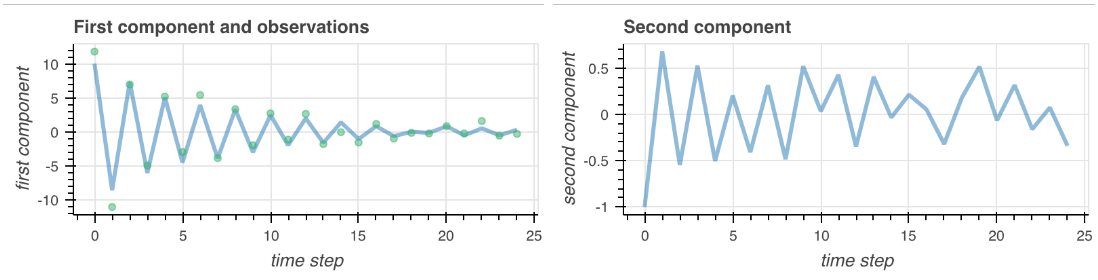
We therefore have the following properties in the complex case:

Complex eigenvalues with $d = 2$ If the transition matrix F has two complex-conjugate eigenvalues λ_1, λ_2 defined as in (3.3) then

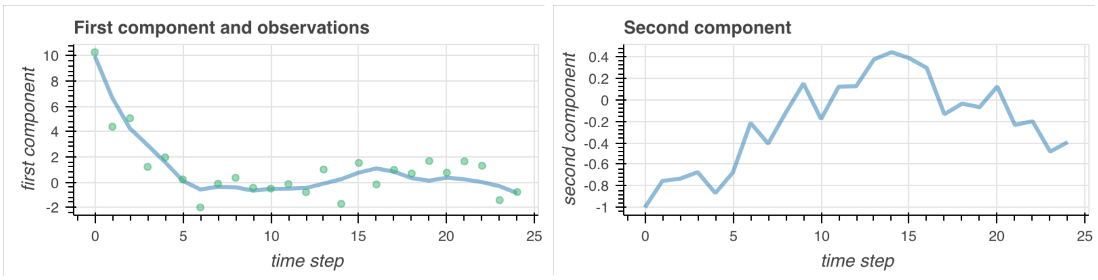
- any model $\mathcal{M}'$ with transition matrix $\lambda R(\omega)$ and observation matrix $(1, 0)$ is said to be a *canonical similar model*



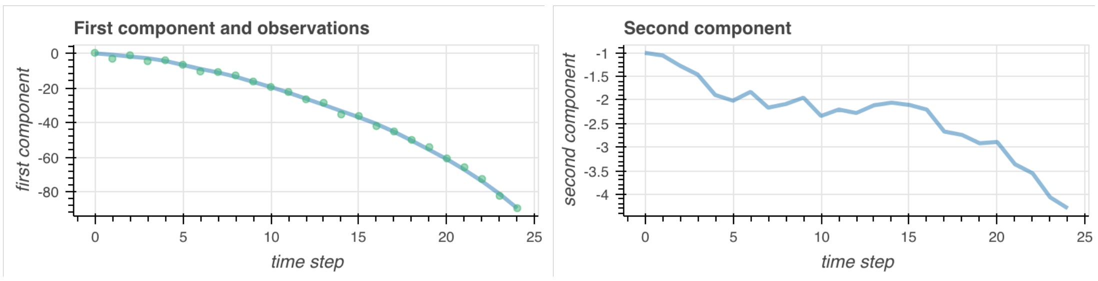
(a) $\lambda = -1.05$, $\theta_0 = (0, -1)^\intercal$



(b) $\lambda = -0.75$, $\theta_0 = (10, -1)^\intercal$

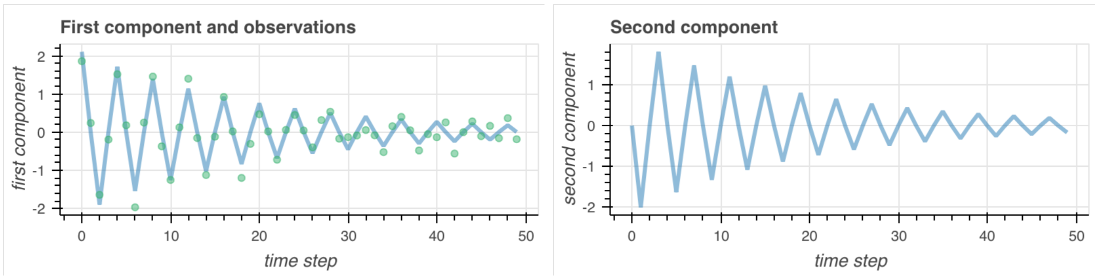


(c) $\lambda = 0.75$, $\theta_0 = (10, -1)^\intercal$

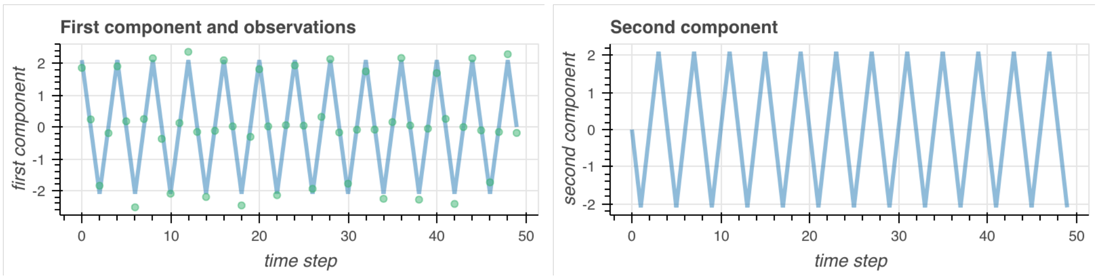


(d) $\lambda = 1.05$, $\theta_0 = (0, -1)^\intercal$

Figure 3.1: The two components and observations of a 2-dimensional DLM with one real eigenvalue λ, initialised at θ_0 and with $U = 0.4 I_2$ and $V = 1$.



(a) $\lambda = 0.95$



(b) $\lambda = 1$

Figure 3.2: The two components and observations of a 2-dimensional DLM with two complex eigenvalues, initialised at θ_0, without evolution noise, with $\theta_0 = (2, 0)^\intercal$ and $\omega = \pi/2$ and with $V = 0.25^2$.

- the canonical similar model $\mathcal{M}^*$ with noise matrices $SU_kS^\intercal$ and V_k, with $S = (T^*)^{-1}T$, and prior $\mathrm{N}(\cdot\,; Sm_0, SP_0S^\intercal)$ is said to be the *canonical equivalent model*.
- the observability matrix T^* of $\mathcal{M}^*$ is

$$T^* = \begin{pmatrix} 1 & 0 \\ \lambda\cos(\omega) & \lambda\sin(\omega) \end{pmatrix}$$

Illustrations of the behaviour of a DLM with two complex eigenvalue are given in Figure 3.2.

3.2 Building DLMs

We will first introduce some simple models in order to build more complex DLMs later in this chapter. These simple models will also illustrate some of the concepts introduced in Section 3.1.

3.2.1 First- and second-order polynomial trend models

A DLM $\mathcal{M}$ in the class of observable time-series DLMs $\mathbb{M}$ is said to be a *d-th order polynomial DLM* if its forecast function is of the form

$$g_k(\delta) = a_{k,0} + a_{k,1}\delta + \cdots + a_{k,d}\delta^{d-1}.$$

We will be focusing on the first- and second-order polynomial DLMs since they are the most useful in practice.

First order

The canonical model for the class of first-order polynomial DLMs is

$$\begin{aligned} \boldsymbol{\theta}_k &= \boldsymbol{\theta}_{k-1} + \boldsymbol{u}_k \\ \boldsymbol{y}_k &= \boldsymbol{\theta}_k + \boldsymbol{v}_k \end{aligned}$$

with $\boldsymbol{u}_k \sim \mathrm{N}(\cdot\,;0,U_k)$ and $\boldsymbol{v}_k \sim \mathrm{N}(\cdot\,;0,V_k)$, that is $F = H = 1$. In spite of its simplicity, this model can be useful when little information is available about the evolution of the quantity of interest. However, this also means that there is no way of learning any complex trend with this model, so that the forecasting capabilities of this model will be limited in general. In particular, the forecast function g_k will be constant and equal to $g_k(0) = \hat{m}_k$.

Second order

In order to specify the form of the canonical model for the class of second-order polynomial DLMs, we consider $\boldsymbol{\theta}_k = (\boldsymbol{x}_k, \dot{\boldsymbol{x}}_k)^\intercal$ with $\boldsymbol{x}_k$ and $\dot{\boldsymbol{x}}_k$ two real-valued random variables (the reason for the notation $\dot{\boldsymbol{x}}_k$ will become clear shortly). The state and observation equations of the associated canonical model are

$$\begin{aligned}
\boldsymbol{x}_k &= \boldsymbol{x}_{k-1} + \dot{\boldsymbol{x}}_{k-1} + \boldsymbol{u}_{k,1} \\
\dot{\boldsymbol{x}}_k &= \dot{\boldsymbol{x}}_{k-1} + \boldsymbol{u}_{k,2} \\
\boldsymbol{y}_k &= \boldsymbol{x}_k + \boldsymbol{v}_k
\end{aligned}$$

with $\boldsymbol{u}_k = (\boldsymbol{u}_{k,1}, \boldsymbol{u}_{k,2})^\intercal \sim \mathrm{N}(\cdot\,;0,U_k)$ and $\boldsymbol{v}_k \sim \mathrm{N}(\cdot\,;0,V_k)$, that is

$$F = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad \text{and} \qquad H = \begin{pmatrix} 1 & 0 \end{pmatrix}.$$

The first component $\boldsymbol{x}_k$ of the parameter $\boldsymbol{\theta}_k$ can be interpreted as the *level* or as the *position* depending on the application while the component $\dot{\boldsymbol{x}}_k$ can be seen as the growth or as the velocity. It can be easily verified that

$$F^\delta = \begin{pmatrix} 1 & \delta \\ 0 & 1 \end{pmatrix}$$

for any $\delta \geq 0$ so that the forecast function can be expressed as

$$\begin{aligned}
g_k(\delta) &= HF^\delta \mathbb{E}(\boldsymbol{\theta}_k \,|\, \boldsymbol{y}_{0:k} = y_{0:k}) \\
&= \mathbb{E}(\boldsymbol{x}_k \,|\, \boldsymbol{y}_{0:k} = y_{0:k}) + \delta \mathbb{E}(\dot{\boldsymbol{x}}_k \,|\, \boldsymbol{y}_{0:k} = y_{0:k}).
\end{aligned}$$

Assuming that the actual evolution of the quantity of interest corresponds to a nearly-constant growth, a second-order polynomial DLM will provide better forecasting capabilities than the first-order one. Although polynomial DLMs of order 3 and above would, in theory, further improve forecasting capabilities, they are rarely used in practice since the higher-order components might be difficult to learn. For instance, a third-order polynomial DLM would correspond to a nearly-constant acceleration model.

Example 3.2. In previous examples, the nearly-constant velocity model was introduced with a transition matrix equal to

$$F = \begin{pmatrix} 1 & \Delta \\ 0 & 1 \end{pmatrix}$$

with Δ the duration of a time step. To verify that this model is similar to a second-order polynomial DLM, we can express it in a canonical form by first computing the eigenvalues via the characteristic polynomial of F:

$$|F - \lambda I_2| = (1 - \lambda)^2$$

which has root $\lambda = 1$ with multiplicity 2. It follows that F is indeed similar to $J_2(1)$ which corresponds to a second order polynomial DLM. To help with the interpretation of this result, we compute the corresponding similarity matrix as

$$S = \begin{pmatrix} 1 & 0 \\ 0 & \Delta \end{pmatrix}$$

which leads to a state $\boldsymbol{\theta}^*_k$ in the canonical equivalent model of the form

$$\boldsymbol{\theta}^*_k = \begin{pmatrix} \boldsymbol{\theta}_{k,1} \\ \Delta \boldsymbol{\theta}_{k,2} \end{pmatrix}.$$

Indeed, a growth rate of $\boldsymbol{\theta}_{k,2}$ over a period Δ is equivalent to a growth rate of $\Delta \boldsymbol{\theta}_{k,2}$ over a period 1.

3.2.2 Seasonal models

Although, in general, one cannot expect to have reliable forecasts over long periods of time when the evolution of the parameter $\boldsymbol{\theta}_k$ is far from polynomial, there exists a special case of interest appearing in many applications, that is when the changes in $\boldsymbol{\theta}_k$ tend to repeat themselves over some given intervals. For instance, this is often the case with data collected over several days, weeks or years which might contain changes that are consistent with day/night, weekday/weekend, or seasons. This is why the corresponding models are often referred to as *seasonal* models. This class of models can be described in generality via a *form-free* approach. However, it can be advantageous in some cases to rely on more restrictive assumptions based on trigonometric functions. The corresponding approach will be called a *Fourier-form* representation of seasonality. Both cases will rely on the following concepts, defined for any function $f : \mathbb{N}_0 \to \mathbb{R}$

1. the function f is *cyclical* or *periodic* if for some integer $p > 1$ it holds that

$$f(k + np) = f(k) \qquad \text{for all integers } k, n \geq 0 \tag{3.7}$$

2. the smallest integer p such that (3.7) holds is called the *period* of f
3. a *full cycle* is any interval of $\mathbb{N}_0$ containing p points, e.g., $[k, k + p - 1]$ for some $k \geq 0$
4. the set of *seasonal factors* is $\{f(k) : 0 \leq k < p\}$, these values are then repeated over any full cycle
5. the *seasonal factor vector* at time step $k \geq 0$ is

$$\theta_k = (f(k), f(k + 1), \ldots, f(k + p - 1))^{\intercal}$$

Form free

If f is a cyclical function then the consecutive seasonal factor vectors θ_{k-1} and θ_k are related by $\theta_k = P\theta_{k-1}$ with P the $p \times p$ matrix of the form

$$P = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & 1 \\ 1 & 0 & 0 & \ldots & 0 \end{pmatrix}$$

The matrix P is called a *permutation* matrix since it corresponds to the permutation σ of $\{1, \ldots, p\}$ such that

$$\sigma(p) = 1 \qquad \text{and} \qquad \sigma(i) = i + 1 \quad \text{for any } i \in \{1, \ldots, p - 1\}.$$

A canonical form-free seasonal factor DLM $\mathcal{M}$ of period p can then be defined as a noisy version of θ_k, that is as

$$\begin{aligned} \boldsymbol{\theta}_k &= P\boldsymbol{\theta}_{k-1} + \boldsymbol{u}_k \\ \boldsymbol{y}_k &= e_p\boldsymbol{\theta}_k + \boldsymbol{v}_k \end{aligned}$$

for some $\boldsymbol{u}_k \sim \mathrm{N}(\cdot\,; 0, U_k)$ and $\boldsymbol{v}_k \sim \mathrm{N}(\cdot\,; 0, V_k)$. This model has very specific properties:

- the DLM $\mathcal{M}$ is observable with observability matrix $T = I_p$
- for some $\delta \geq 0$, let $i \equiv \delta \pmod p$, then the forecast function verifies

$$g_k(\delta) = e_p P^{\delta} \mathbb{E}(\boldsymbol{\theta}_k \,|\, \boldsymbol{y}_{0:k} = y_{0:k}) = \mathbb{E}(\boldsymbol{\theta}_{k,i+1} \,|\, \boldsymbol{y}_{0:k} = y_{0:k})$$

Fourier form

When appropriate, the use of trigonometric functions can be highly beneficial since they provide a simple parametrisation of the evolution of $\boldsymbol{\theta}_k$ and can often be simpler to interpret. For instance, if the evolution of $\boldsymbol{\theta}_k$ can be faithfully modelled by a single sine or cosine then only two parameters are required, i.e. the phase and the amplitude, whereas a form-free approach will require as many seasonal factors as there are time steps within one period.

Any cyclical function $f : \mathbb{N}_0 \to \mathbb{R}$ can be represented on $\{0, \ldots, p-1\}$ as

$$f(j) = a_0 + \sum_{r=1}^{h} \big(a_r \cos(\omega r j) + b_r \sin(\omega r j) \big)$$

with $h = \lfloor p/2 \rfloor$, $\omega = 2\pi/p$ and with

$$a_r = \begin{cases} \dfrac{1}{p} \displaystyle\sum_{j=0}^{p-1} f(j) & \text{if } r = 0 \\ \dfrac{1}{p} \displaystyle\sum_{j=0}^{p-1} (-1)^j f(j) & \text{if } r = p/2 \\ \dfrac{2}{p} \displaystyle\sum_{j=0}^{p-1} f(j) \cos(\omega r j) & \text{otherwise} \end{cases} \quad \text{and} \quad b_r = \begin{cases} 0 & \text{if } r = p/2 \\ \dfrac{2}{p} \displaystyle\sum_{j=0}^{p-1} f(j) \sin(\omega r j) & \text{otherwise} \end{cases}$$

One of the simplest Fourier-form seasonal DLM follows from a forecast function of the following form

$$g_k(\delta) = a_k \cos(\omega\delta) + b_k \sin(\omega\delta) = e_2 R(\omega)^\delta \begin{pmatrix} a_k \\ b_k \end{pmatrix}$$

for some coefficient $a_k, b_k \in \mathbb{R}$. With $\mathbb{E}(\boldsymbol{\theta}_k \,|\, \boldsymbol{y}_{0:k} = y_{0:k}) = (a_k, b_k)^\intercal$, this corresponds to a DLM with observation matrix e_2 and transition matrix $R(\omega)$, which we refer to as a *harmonic* DLM. In the case where the frequency ω is equal to π, this DLM is not observable and reduces to a DLM with the observation and transition matrix 1 and -1 respectively. Otherwise, when $\omega \in (0, \pi)$, the harmonic DLM is observable with observability matrix

$$T = \begin{pmatrix} 1 & 0 \\ \cos(\omega) & \sin(\omega) \end{pmatrix}.$$

3.2.3 Superposition

We now have a few elementary building blocks available for modelling the evolution and observation of the quantities of interest. Yet, it is often the case in real data that several of these models must be combined to adequately fit the available observations and to make reliable forecasts. The following theorem shows that time series that can be expressed as the sum of elementary time series can also be modelled by a DLM which arises as the *superposition* of the corresponding elementary DLMs.

Theorem 3.1. *Let $(\boldsymbol{y}_k)_{k \geq 0}$ be a time series defined as*

$$\boldsymbol{y}_k = \sum_{i=1}^{h} \boldsymbol{y}_{i,k},$$

where h is a positive integer and where, for any $i \in \{1, \ldots, h\}$, $(\boldsymbol{y}_{i,k})_{k \geq 0}$ is a time series modelled by a DLM $\mathcal{M}_i$ of dimension d_i whose components are denoted $\boldsymbol{\theta}_{i,k}$, $F_{i,k}$, $H_{i,k}$, $U_{i,k}$ and $V_{i,k}$. If, for any DLM $\mathcal{M}_i$, the corresponding noise vectors $\boldsymbol{u}_{i,k}$ and $\boldsymbol{v}_{i,k}$ are mutually independent of $\boldsymbol{u}_{j,k}$ and $\boldsymbol{v}_{j,k}$, $j \neq i$, then the time series $(\boldsymbol{y}_k)_{k \geq 0}$ is also modelled by a DLM of dimension $d = d_1 + \cdots + d_h$ characterised by

$$\boldsymbol{\theta}_k = \begin{pmatrix} \boldsymbol{\theta}_{1,k} \\ \vdots \\ \boldsymbol{\theta}_{h,k} \end{pmatrix}, \qquad F_k = \begin{pmatrix} F_{1,k} & & \\ & \ddots & \\ & & F_{h,k} \end{pmatrix}, \qquad U_k = \begin{pmatrix} U_{1,k} & & \\ & \ddots & \\ & & U_{h,k} \end{pmatrix}$$

as well as

$$H_k = \begin{pmatrix} H_{1,k} & \ldots & H_{h,k} \end{pmatrix} \qquad \text{and} \qquad V_k = \sum_{i=1}^{h} V_{i,k}.$$

3.2.4 Defining the transition noise

One important but difficult task that has not been discussed yet is to define the covariance matrix of the transition noise U_k of a given DLM $\mathcal{M}$. This matrix determines how much the corresponding evolution model can be trusted. For instance, in the one-dimensional case, if the value of U_k is 0 then the transition is deterministic which is equivalent to assuming that the state evolves exactly according to the chosen model. Alternatively, if the value of U_k tends to infinity, then the transition is uninformative and cannot be used for any useful forecasting. In higher-dimensional settings, defining U_k is even more challenging. One simple way to address this difficulty is to rely on the precision matrix arising from a deterministic prediction and apply a discount factor $\alpha \in (0, 1)$ to it in order to model a loss of information. Formally, if we denote $\hat{P}_{k-1} = \mathrm{var}(\boldsymbol{\theta}_{k-1} \,|\, \boldsymbol{y}_{0:k-1} = y_{0:k-1})$ the updated covariance matrix at time step $k - 1 \geq 0$, then the corresponding predicted covariance matrix in a deterministic context if $\tilde{P}_k = F_k \hat{P}_{k-1} F_k^{\intercal}$. The associated precision matrix is simply $\tilde{P}_k^{-1}$ and the discounted version of it is $\alpha \tilde{P}_k^{-1}$. This approach yields a predicted covariance matrix P_k expressed as

$$P_k = \mathrm{var}(\boldsymbol{\theta}_k \,|\, \boldsymbol{y}_{0:k-1} = y_{0:k-1}) = \frac{1}{\alpha} \tilde{P}_k.$$

In order to identify the value of U_k corresponding to this approach, remember that $P_k = \tilde{P}_k + U_k$, so that

$$U_k = \frac{1-\alpha}{\alpha} \tilde{P}_k = \frac{1-\alpha}{\alpha} F_k \hat{P}_{k-1} F_k^{\intercal}.$$

Note that the value of α is typically set in the interval $[0.9, 0.99]$. In the case of a superposed model, the different elementary models often require distinct discount factor, which can be easily implemented.

In some cases, it is possible to determine the matrix U_k from physical principles. For instance, in the situation where $\boldsymbol{\theta}_k = (\boldsymbol{x}_k, \dot{\boldsymbol{x}}_k)^{\intercal}$ with $\boldsymbol{x}_k$ and $\dot{\boldsymbol{x}}_k$ representing position and velocity respectively, one can that the velocity will be affected by a noise with standard deviation $a\Delta$ (the integral of the constant acceleration a over one time step of duration Δ) and the position will be affected by a noise with standard deviation $a\Delta^2/2$ (again by integration). It follows that the state equation can be expressed as

$$\boldsymbol{\theta}_k = \begin{pmatrix} 1 & \Delta \\ 0 & 1 \end{pmatrix} \boldsymbol{\theta}_{k-1} + a \begin{pmatrix} \Delta^2/2 \\ \Delta \end{pmatrix} \boldsymbol{\epsilon}_k,$$

where a is the standard deviation of the random acceleration (zero-mean) and where $(\boldsymbol{\epsilon}_k)_k$ is an i.i.d. sequence of random variables with standard normal distribution. This can be expressed equivalently as

$$\boldsymbol{\theta}_k = \begin{pmatrix} 1 & \Delta \\ 0 & 1 \end{pmatrix} \boldsymbol{\theta}_{k-1} + \boldsymbol{u}_k$$

with $\boldsymbol{u}_k \sim \mathrm{N}(\cdot\,; 0, U)$ where

$$U = a^2 \begin{pmatrix} \Delta^2/2 \\ \Delta \end{pmatrix} \begin{pmatrix} \Delta^2/2 \\ \Delta \end{pmatrix}^{\intercal} = a^2 \begin{pmatrix} \Delta^4/4 & \Delta^3/2 \\ \Delta^3/2 & \Delta^2 \end{pmatrix}.$$

There are often several ways of defining the covariance matrix of the transition noise. For instance, one could alternatively derive the form of U from a continuous-time nearly-constant velocity model. This would yield a different expression

$$U = a' \begin{pmatrix} \Delta^3/3 & \Delta^2/2 \\ \Delta^2/2 & \Delta \end{pmatrix},$$

with $a' > 0$ another way of characterising the noise on the acceleration.