Dynamic Linear Models

January 21, 2021

We call **information set** at time t the set of all the **known** informations up to time t, and we denote it by $\mathcal{D}_t$.

Starting examples

Example 1. *First-order polynomial model*

Given a time series Y_t, we model the series as follow

$$Y_t = \mu_t + v_t \qquad v_t \sim \mathcal{N}(0, \sigma_t^2)$$

the term v_t is named **observational error**. Here μ_t is modeling the level of the time series at time t. We further model how the level changes in time as a random walk

$$\mu_t = \mu_{t-1} + \omega_t \qquad \omega_t \sim N(0, w_t)$$

such dynamic is also named **locally constant mean**, since the dynamic of μ_t is slow. Indeed at each time step t μ_t is equal to the value assumed at the time step before μ_t plus a small variation centered in 0. Hence a random walk generally identfies a slow variation dynamic, if the **evolutional error** variance w_t is small.

We also assume $v_t \perp v_s$, $\omega_t \perp \omega_s$, $v_t \perp \omega_s$ for all time instant $t \neq s$. Morover at this stage we assume both variances σ_t^2 and w_t known for every time instant t. Given all the previous assumptions the dynamic model is so formulated:

$$\begin{cases} Y_t = \mu_t + v_t & v_t \sim \mathcal{N}(0, \sigma_t^2) \\ \mu_t = \mu_{t-1} + \omega_t & \omega_t \sim \mathcal{N}(0, w_t) \end{cases}$$

More compactly we can write

$$\begin{cases} Y_t|\mu_t & \sim \mathcal{N}(\mu_t, \sigma_t^2) \\ \mu_t|\mu_{t-1} & \sim \mathcal{N}(\mu_{t-1}, w_t) \end{cases}$$

About the meaning of this model, we are saying that the level of the series is locally constant in time, since μ_t follows a random walk dynamic. Hence we expect significant changes of the trend in the long time, but this first order polynomial model do not model such long time changes.

An interesting quantity of interest to understand this model is the **expected level of the series k step ahead** conditional on the current level μ_t of the series

$$\mathbb{E}[y_{t+k}|\mu_t] = \mathbb{E}[\mu_{t+k} + v_t|\mu_t] = [\mu_{t+k}|\mu_t] = \mathbb{E}\left[\mu_t + \sum_{i=1}^{k} \omega_{t+i} \middle| \mu_t\right] = \mu_t + \sum_{i=1}^{k} \mathbb{E}[\omega_{t+i}|\mu_t] = \mu_t$$

From here the name **locally constant mean** model, given the value of the level now we expect that for any time step ahead k it will not change.

Definition 1. *Given information available at time t, represented by the information set $\mathcal{D}_t$, we call **forecast function** the expected value of the posterior distribution of Y_t k step ahead*

$$f_t(k) = \mathbb{E}[Y_{t+k}|\mathcal{D}_t]$$

For the first order polynomial model we have that the forecast function is constant, being given by:

$$f_t(k) = \mathbb{E}[y_{t+k}|\mathcal{D}_t] = \mathbb{E}[\mu_t|\mathcal{D}_t] = m_t \qquad \forall k > 0$$

Consequently, this DLM is useful only for short-term forecasting.

The core of all the DLM theory (once normally distributed errors are assumed) is in the recurent nature of the update parameter equations. A dynamic linear model is defined once an intial condition on μ_0 is set. We assume

$$\mu_0|\mathcal{D}_0 \sim \mathcal{N}(m_0, C_0) \qquad v_t, \omega_t \perp \mu_0|\mathcal{D}_0 \quad \forall t$$

Definition 2. *For each t, the first order polynomial DLM is defined by*

$$\begin{cases} Y_t = \mu_t + v_t \quad v_t \sim \mathcal{N}(0, \sigma_t^2) & \textit{Observation equation} \\ \mu_t = \mu_{t-1} + \omega_t \quad \omega_t \sim \mathcal{N}(0, w_t) & \textit{System equation} \\ \mu_0|\mathcal{D}_0 \sim \mathcal{N}(m_o, C_0) & \textit{Initial information} \end{cases}$$

where the error sequences v_t and ω_t are internally independent and mutually independent, and independent of intial information $\mu_0|\mathcal{D}_0$.

Morover we assume that each information set $\mathcal{D}_t$ at any time t contains: the initial information $\mathcal{D}_0$, the variance values σ_t^2 and w_t for all times $t > 0$ (we are assuming the variances to be known) and past observations $\{y_1, y_2, \ldots, y_{t-1}, y_t\}$. Doing so, the only new information we collect at time $t+1$ is the observed process value y_{t+1}, so that

$$\mathcal{D}_{t+1} = \mathcal{D}_t \cup \{y_{t+1}\}$$

Theorem 1. ***(Updating equations)****: Given the dynamic linear model*

$$\begin{cases} Y_t = \mu_t + v_t \quad v_t \sim \mathcal{N}(0, \sigma_t^2) & \textit{Observation equation} \\ \mu_t = \mu_{t-1} + \omega_t \quad \omega_t \sim \mathcal{N}(0, w_t) & \textit{System equation} \\ \mu_0|\mathcal{D}_0 \sim \mathcal{N}(m_o, C_0) & \textit{Initial information} \end{cases}$$

for any time $t > 0$ the one step forecast and posterior distributions can be obtained sequentially as:

$$\begin{array}{lrl} \mu_t|\mathcal{D}_{t-1} \sim \mathcal{N}(m_{t-1}, R_t) & R_t = C_{t-1} + w_t & \textit{Prior for } \mu_t \\ Y_t|\mathcal{D}_{t-1} \sim \mathcal{N}(f_t, Q_t) & Q_t = R_t + \sigma_t^2 & \textit{One step forecast} \\ & f_t = m_{t-1} & \\ \mu_t|\mathcal{D}_t \sim \mathcal{N}(m_t, C_t) & m_t = m_{t-1} + A_t e_t \quad C_t = A_t \sigma_t^2 & \textit{Posterior for } \mu_t \\ & A_t = R_t Q_t^{-1} \quad e_t = Y_t - f_t & \end{array}$$

Proof: By induction, let $\mu_{t-1}|\mathcal{D}_{t-1} \sim \mathcal{N}(m_{t-1}, C_{t-1})$. This is true for $t = 1$ by assumption (see intitial condition). Thus, from the system equation and the fact that the sum of two indipendent normal random variables is still a normal random variable we have:

$$\begin{cases} \mu_{t-1}|\mathcal{D}_{t-1} & \sim \mathcal{N}(m_{t-1}, C_{t-1}) \\ \omega_t|\mathcal{D}_{t-1} & \sim \mathcal{N}(0, w_t) \end{cases} \qquad \mu_{t-1}|\mathcal{D}_{t-1} \perp \omega_t|\mathcal{D}_{t-1} \qquad \to \qquad \mu_t = \mu_{t-1} + \omega_t \sim \mathcal{N}(m_{t-1}, C_{t-1} + w_t)$$

In the same way we derive the one step forecast distribution. Let $R_t = C_{t-1} + w_t$, starting from the observation equation we obtain:

$$\begin{cases} \mu_t|\mathcal{D}_{t-1} & \sim \mathcal{N}(m_{t-1}, R_t) \\ v_t|\mathcal{D}_{t-1} & \sim \mathcal{N}(0, \sigma_t^2) \end{cases} \qquad \mu_t|\mathcal{D}_{t-1} \perp v_t|\mathcal{D}_{t-1} \qquad \to \qquad Y_t = \mu_t + v_t \sim \mathcal{N}(m_{t-1}, R_t + \sigma_t^2)$$

Let now $\mathcal{D}_t = \mathcal{D}_{t-1} \cup \{y_t\}$, to derive the posterior for μ_t given $\mathcal{D}_t$ we apply bayes' theorem

$$\pi(\mu_t|\mathcal{D}_t) \propto f(Y_t|\mathcal{D}_{t-1}, \mu_t)\pi(\mu_t, \mathcal{D}_{t-1})$$

because we are conditioning Y_t with respect to μ_t, from observation equation it directly follows

$$Y_t|\mathcal{D}_{t-1}, \mu_t \sim \mathcal{N}(\mu_t, \sigma_t^2)$$

so that

$$\begin{aligned}
f(Y_t|\mathcal{D}_{t-1}, \mu_t)\pi(\mu_t, \mathcal{D}_{t-1}) &\propto \exp\left\{-\frac{1}{2\sigma_t^2}(Y_t - \mu_t)^2 - \frac{1}{2R_t}(\mu_t - m_{t-1})^2\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma_t^2}(Y_t^2 + \mu_t^2 - 2Y_t\mu_t) - \frac{1}{2R_t}(\mu_t^2 + m_{t-1}^2 - 2\mu_t m_{t-1})\right\} \\
&\propto \exp\left\{-\left(\frac{1}{2\sigma_t^2} + \frac{1}{2R_t}\right)\mu_t^2 + 2\left(\frac{Y_t}{2\sigma_t^2} + \frac{m_{t-1}}{2R_t}\right)\mu_t - \left(\frac{Y_t^2}{2\sigma_t^2} + \frac{m_{t-1}^2}{2R_t}\right)\right\} \\
&\propto \exp\left\{-\left(\frac{1}{2\sigma_t^2} + \frac{1}{2R_t}\right)\mu_t^2 + 2\left(\frac{Y_t}{2\sigma_t^2} + \frac{m_{t-1}}{2R_t}\right)\mu_t\right\} \\
&= \exp\left\{-\left(\frac{R_t + \sigma_t^2}{2\sigma_t^2 R_t}\mu_t^2 - 2\frac{R_tY_t + \sigma_t^2 m_{t-1}}{2\sigma_t^2 R_t}\mu_t\right)\right\} \\
&= \exp\left\{-\frac{R_t + \sigma_t^2}{2\sigma_t^2 R_t}\left(\mu_t^2 - 2\frac{R_tY_t + \sigma_t^2 m_{t-1}}{R_t + \sigma_t^2}\mu_t + \left[\frac{R_tY_t + \sigma_t^2 m_{t-1}}{R_t + \sigma_t^2}\right]^2 - \left[\frac{R_tY_t + \sigma_t^2 m_{t-1}}{R_t + \sigma_t^2}\right]^2\right)\right\} \\
&\propto \exp\left\{-\frac{R_t + \sigma_t^2}{2\sigma_t^2 R_t}\left(\mu_t - \frac{R_tY_t + \sigma_t^2 m_{t-1}}{R_t + \sigma_t^2}\right)^2\right\}
\end{aligned}$$

We have obtained the kernel of a gaussian distribution of variance

$$C_t = \frac{\sigma_t^2 R_t}{R_t + \sigma_t^2} = \frac{R_t}{Q_t}\sigma_t^2 = A_t\sigma_t^2 \qquad A_t = \frac{R_t}{Q_t}$$

and mean

$$\begin{aligned}
m_t &= \frac{R_tY_t + \sigma_t^2 m_{t-1}}{R_t + \sigma_t^2} = C_t\frac{R_tY_t + \sigma_t^2 m_{t-1}}{R_t\sigma_t^2} = A_t\sigma_t^2\left(\frac{Y_t}{\sigma_t^2} + \frac{m_{t-1}}{R_t}\right) = A_tY_t + \frac{A_t\sigma_t^2}{R_t}m_{t-1} = \\
&= A_tY_t + \frac{\sigma_t^2}{Q_t}m_{t-1} = A_tY_t + \frac{\sigma_t^2}{Q_t}m_{t-1} + \frac{R_t}{Q_t}m_{t-1} - \frac{R_t}{Q_t}m_{t-1} = A_tY_t + m_{t-1} - \frac{R_t}{Q_t}m_{t-1} = \\
&= A_tY_t + m_{t-1} - A_tm_{t-1} = m_{t-1} + A_t(Y_t - m_{t-1}) = m_{t-1} + A_te_t
\end{aligned}$$

where $e_t = Y_t - m_{t-1}$ is the **one step ahead forecast error**, and is the difference between the observed value Y_t and the estimated one m_{t-1}. The term A_t is named instead **adaptive coefficient**.

The previous result will be proved in its generality for any DLM. Anyway the key result reside in the sequential update of forecast and posterior distributions: given a new observation Y_t current estimate of the system state is updated so to correct with respect to the error committed. In this way the model is able to track the time series as new information arrives.

Theorem 2. *For $k > 0$, the k-step ahead forecast distribution conditioned on information available at time t is*

$$Y_{t+k}|\mathcal{D}_t \sim \mathcal{N}(m_t, Q_t(k))$$

where

$$Q_t(k) = C_t + \sum_{i=1}^{k} w_{t+i} + \sigma_{t+k}^2$$

Proof: From the evolution equation for μ_t and the observation equation for Y_t, for $k \geq 1$, we have

$$Y_{t+k} = \mu_{t+k} + v_{t+k} = \left(\mu_t + \sum_{i=1}^{k} \omega_{t+i} \right) + v_{t+k}$$

Since all terms are normal and mutually independent once conditioned on $\mathcal{D}_t$, the result follow.

We now consider the case where the observational noise variance σ_t^2 is constant and equal to σ^2 but **unknown**. In this case we need to put prior distribution also on σ^2. Generally the standard normal-inverse gamma model is assumed. This structure enables a conjugate sequential updating procedure for σ^2, in addition to that for μ_t.

Definition 3. *For each t, the first order polynomial DLM is defined by*

$$\begin{cases} Y_t = \mu_t + v_t \quad v_t \sim \mathcal{N}(0, \sigma^2) & \textit{Observation equation} \\ \mu_t = \mu_{t-1} + \omega_t \quad \omega_t \sim \mathcal{N}(0, \sigma^2 w_t) & \textit{System equation} \\ \mu_0 | \sigma^2, \mathcal{D}_0 \sim \mathcal{N}(m_o, \sigma^2 C_0) & \textit{Initial information} \\ \sigma^2 | \mathcal{D}_0 \sim inv\Gamma\left(\frac{n_0}{2}, \frac{d_0}{2}\right) \end{cases}$$

for some known m_0, C_0, w_t, n_0 and d_0. The error sequences v_t and ω_t are internally independent and mutually independent, and independent of intial information $\mu_0|\mathcal{D}_0$, once conditioned on σ^2.

It can be shown with similar computations as for the case with known observational variance that, given initial conditions as in definition 3, the following holds:

Theorem 3. ***(Updating equations)****: Given the dynamic linear model*

$$\begin{cases} Y_t = \mu_t + v_t \quad v_t \sim \mathcal{N}(0, \sigma^2) & \textit{Observation equation} \\ \mu_t = \mu_{t-1} + \omega_t \quad \omega_t \sim \mathcal{N}(0, \sigma^2 w_t) & \textit{System equation} \\ \mu_0 | \sigma^2, \mathcal{D}_0 \sim \mathcal{N}(m_o, \sigma^2 C_0) & \textit{Initial information} \\ \sigma^2 | \mathcal{D}_0 \sim inv\Gamma\left(\frac{n_0}{2}, \frac{d_0}{2}\right) \end{cases}$$

for any time $t > 0$ the one step forecast and posterior distributions conditioned on the observational variance σ^2 can be obtained sequentially as:

$$\begin{array}{lrl} \sigma^2 | \mathcal{D}_{t-1} \sim inv\Gamma\left(\dfrac{n_{t-1}}{2}, \dfrac{d_{t-1}}{2}\right) & & \textit{Prior for } \sigma^2 \\ \mu_t | \mathcal{D}_{t-1}, \sigma^2 \sim \mathcal{N}(m_{t-1}, \sigma^2 R_t) & R_t = C_{t-1} + w_t & \textit{Prior for } \mu_t \\ Y_t | \mathcal{D}_{t-1}, \sigma^2 \sim \mathcal{N}(f_t, \sigma^2 Q_t) & Q_t = R_t + 1 & \textit{One step forecast} \\ & f_t = m_{t-1} & \\ \sigma^2 | \mathcal{D}_t \sim inv\Gamma\left(\dfrac{n_t}{2}, \dfrac{d_t}{2}\right) & n_t = n_{t-1} + 1 & \textit{Posterior for } \sigma^2 \\ & d_t = d_{t-1} + \dfrac{e_t^2}{Q_t} & \\ \mu_t | \mathcal{D}_t, \sigma^2 \sim \mathcal{N}(m_t, \sigma^2 C_t) & m_t = m_{t-1} + A_t e_t \quad C_t = A_t & \textit{Posterior for } \mu_t \\ & A_t = R_t Q_t^{-1} \quad e_t = Y_t - f_t & \end{array}$$

Note that the model is conjugate and that the distributions are given conditionally on σ^2.

Example 2. *Dynamic regression model*

Regression modeling for a time series concerns in relating the mean response function μ_t of the original series to another time series of observations X_t via a particular regression function. For example, let X_t a regressor time series, a simple linear model for the effect of X_t on the current mean of the target series Y_t is

$$\mu_t = \alpha_t + \beta_t X_t$$

If we believe that the real mean response function is sufficiently smooth and well-behaved locally as a function of both X_t and t, then for short-term prediction, that is, local inference, a local approximating model of the form $\mu_t = \alpha_t + \beta_t X_t$ may be enought to forecast the target series Y_t. Suppose that the series we want to forecast is as the one reported in figure 1. Is clear the importance to make the parameters α and β time varing. Indeed in region 1 β must be negative, while in region 2 it must be positive. A static model of the form $\mu_t = \alpha + \beta X_t$ is not able to efficiently track μ_t as function of time.

For this reason we need to introduce some kind of dynamic fo the model parameters. If we believe that parameters changes slowly in time, a random walk dynamic is ok for their description, namely we assume

$$\alpha_t = \alpha_{t-1} + \omega_t^\alpha \qquad \omega_t^\alpha \sim \mathcal{N}(0, (w_t^\alpha)^2)$$
$$\beta_t = \beta_{t-1} + \omega_t^\beta \qquad \omega_t^\beta \sim \mathcal{N}(0, (w_t^\beta)^2)$$

The model can be then written as follow:

$$\begin{cases} \mu_t = \alpha_t + \beta_t X_t \\ \alpha_t = \alpha_{t-1} + \omega_t^\alpha & \omega_t^\alpha \sim \mathcal{N}(0, (w_t^\alpha)^2) \\ \beta_t = \beta_{t-1} + \omega_t^\beta & \omega_t^\beta \sim \mathcal{N}(0, (w_t^\beta)^2) \end{cases}$$

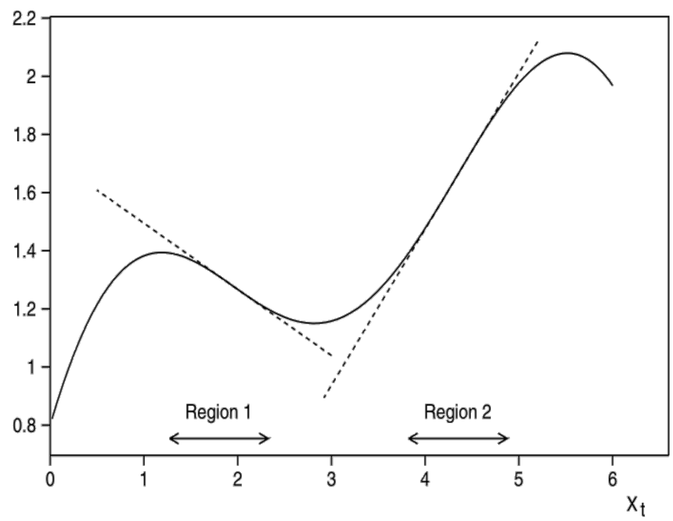

Figure 1: Local linearity of μ_t as a function of X_t

This model is also named **simple regression DLM**, because the response is written as function of just one covariate. The state evolution vector $\boldsymbol{\omega_t}$ is a zero mean random vector. This expresses the concept of local constancy of the parameters, subject to variation controlled by the variance matrix of $\boldsymbol{\omega_t}$: little values of the variance matrix W_t translate a stable and slow evolution of the parameters α_t and β_t, since for any time t, they assume the value assumed at the time step before $t-1$ plus a small zero mean variation. On the contrary big values in W_t imply great variability in the development of μ_t.

Note that this kind of slow variation dynamic does not work at all if there happens to be a large, abrupt change in X_t (some peaks in the time series of X_t). In this case the model parameters are not able to react fast to the change in X_t and hence we are not able to track the target μ_t.

Definition 4. ***Dynamic regression**: Let $\boldsymbol{F_t} = [X_{1,t}, X_{2,t}, \ldots, X_{n,t}]$ a vector of known time series for any time $t > 0$, named regression vector. To include the intercept in the model we assume $X_{1,t} = 1 \quad \forall t$. Let $\boldsymbol{\theta_t} = [\theta_{1,t}\theta_{2,t} \ldots \theta_{n,t}]^T$ the parameter vector at time t. The linear regression DLM is defined by*

$$\begin{cases} Y_t = \boldsymbol{F_t}\boldsymbol{\theta_t} + v_t \quad v_t \sim \mathcal{N}(0, \sigma_t^2) & \textit{Observation equation} \\ \boldsymbol{\theta_t} = \boldsymbol{\theta_{t-1}} + \boldsymbol{\omega_t} \quad \omega_t \sim \mathcal{N}(0, \boldsymbol{W_t}) & \textit{System equation} \\ \boldsymbol{\theta_0}|\mathcal{D}_0 \sim \mathcal{N}(m_0, C_0) & \textit{Initial information} \end{cases}$$

Morover the error sequences v_t and $\boldsymbol{\theta_t}$ are assumed to be internally and mutually indipendent, and indipendent from initial information $\mathcal{D}_0$. $\boldsymbol{W_t}$ is named evolution variance matrix.

The sequential model description for the series requires that the defining quantities at time t be known at that time. Similarly, when forecasting more than one step ahead, let's say up to time $t + k$, given we are at time t, the corresponding quantities F_{t+k} , V_{t+k} , and W_{t+k} must belong to the current information set $\mathcal{D}_t$. (???????)

As for the first order polynomial model, and as will be proved in its generality, the main point of any DLM is in the updating parameter equations. The following result holds for the particular case of simple regression DLM

Theorem 4. ***(Updating equations)**: Given the dynamic linear model*

$$\begin{cases} Y_t = \theta_t X_t + v_t \quad v_t \sim \mathcal{N}(0, \sigma_t^2) & \textit{Observation equation} \\ \theta_t = \theta_{t-1} + \omega_t \quad \omega_t \sim \mathcal{N}(0, w_t) & \textit{System equation} \\ \theta_0 | \mathcal{D}_0 \sim \mathcal{N}(m_o, C_0) & \textit{Initial information} \end{cases}$$

for any time $t > 0$ the one step forecast and posterior distributions can be obtained sequentially as:

$$\theta_t | \mathcal{D}_{t-1} \sim \mathcal{N}(m_{t-1}, R_t) \qquad R_t = C_{t-1} + w_t^2 \qquad \textit{Prior for } \theta_t$$

$$Y_t | \mathcal{D}_{t-1} \sim \mathcal{N}(f_t, Q_t) \qquad Q_t = X_t^2 R_t + \sigma_t^2 \qquad \textit{One step forecast}$$

$$f_t = X_t m_{t-1}$$

$$\theta_t | \mathcal{D}_t \sim \mathcal{N}(m_t, C_t) \qquad m_t = m_{t-1} + A_t e_t \quad C_t = \frac{R_t}{Q_t}\sigma_t^2 \qquad \textit{Posterior for } \theta_t$$

$$A_t = \frac{R_t X_t}{Q_t} \quad e_t = Y_t - f_t$$

Proof: The theorem can be proved using bayes' theorem as done for the first order polynomial model. Anyway we can exploit the gaussian structure of DLM. Thus before proceed some preliminary results about normal distributions:

1. let $\boldsymbol{X} \sim \mathcal{N}_k(\boldsymbol{\mu}, \Sigma)$ a gaussian random vector, given A $k \times k$ matrix

   $$\boldsymbol{Y} = A\boldsymbol{X} \sim \mathcal{N}_k(A\boldsymbol{\mu}, A\Sigma A^T)$$

2. let $\boldsymbol{X} \sim \mathcal{N}_k(\boldsymbol{\mu}, \Sigma)$ and $\boldsymbol{Y} \sim \mathcal{N}_k(\boldsymbol{\gamma}, W)$ two independent gaussian random vectors then

   $$\boldsymbol{X} + \boldsymbol{Y} \sim \mathcal{N}_k(\boldsymbol{\mu} + \boldsymbol{\gamma}, \Sigma + W)$$

The idea of the proof is to build the random vector

$$\begin{bmatrix} \theta_t \\ Y_t \end{bmatrix} | \mathcal{D}_{t-1}$$

show that it is indeed a gaussian random vector and thus computing the posterior distribution of θ_t by computing the conditional distribution $\theta_t | \mathcal{D}_{t-1}, Y_t = \theta_t | \mathcal{D}_t$ which coincides with the posterior for θ_t.

By induction, let assume that $\theta_{t-1} | \mathcal{D}_{t-1} \sim \mathcal{N}(m_{t-1}, C_{t-1})$. This is true for $t = 1$ by assumption (see initial condition). From the state equation we have

$$\theta_t = \theta_{t-1} + \omega_t \quad \rightarrow \quad \theta_t | \mathcal{D}_{t-1} \sim \mathcal{N}(m_{t-1}, C_{t-1} + w_t^2)$$

beeing both θ_{t-1} and ω_t normally distributed and independent. In a similar way, from the observation equation we have

$$Y_t = X_t \theta_t + v_t \quad \rightarrow \quad Y_t | \mathcal{D}_{t-1} \sim \mathcal{N}(X_t m_{t-1}, X_t^2(C_{t-1} + w_t^2) + \sigma_t^2)$$

having applied property 1 of normal distributions recapped before to $X_t \theta_t$ (notice that for this specific DLM all quantities are scalars). Let now $R_t = C_{t-1} + w_t^2$ and $Q_t = X_t^2(C_{t-1} + w_t^2) + \sigma_t^2 = X_t^2 R_t + \sigma_t^2$. Morover

$$\begin{aligned} COV[\theta_t, Y_t | \mathcal{D}_{t-1}] &= COV[\theta_t, X_t\theta_t + v_t | \mathcal{D}_{t-1}] = COV[\theta_t, X_t\theta_t | \mathcal{D}_{t-1}] + COV[\theta_t, v_t | \mathcal{D}_{t-1}] \\ &\overset{\theta_t \perp v_t}{=} X_t VAR[\theta_t | \mathcal{D}_{t-1}] = X_t R_t \end{aligned}$$

Consider now the following gaussian random vector

$$\begin{bmatrix} \theta_t \\ Y_t \end{bmatrix} | \mathcal{D}_{t-1} \sim \mathcal{N}_2\left(\begin{bmatrix} m_{t-1} \\ X_t m_{t-1} \end{bmatrix}, \begin{bmatrix} R_t & X_t R_t \\ X_t R_t & Q_t \end{bmatrix} \right)$$

then the conditional distribution of θ_t given Y_t is given as

$$\theta_t | Y_t, \mathcal{D}_{t-1} = \theta_t | \mathcal{D}_t \sim \mathcal{N}\left(m_{t-1} + \frac{X_t R_t}{Q_t}(Y_t - X_t m_{t-1}), R_t - \frac{X_t^2 R_t^2}{Q_t} \right)$$

Let $A_t = \frac{X_t R_t}{Q_t}$ and $e_t = Y_t - X_t m_{t-1}$, then

$$m_t = m_{t-1} + A_t e_t$$
$$C_t = R_t - \frac{X_t^2 R_t^2}{Q_t} = \frac{R_t Q_t - X_t^2 R_t}{Q_t} = \frac{R_t(X_t^2 R_t + \sigma_t^2) - X_t^2 R_t^2}{Q_t} = \frac{R_t}{Q_t}\sigma_t^2$$

Note that the posterior mean m_t is obtained by correcting the prior mean m_{t-1} with a term proportional to the forecast error e_t.

The general univariate DLM

The first-order polynomial and simple regression models illustrate many basic concepts and important features of the general class of normal dynamic linear models, referred to as dynamic linear models (DLMs).

The crucial structural property enabling dynamic modeling in DLM is conditional independence: **given the present, the future is independent of the past**. DLM are markovian models, where all the information required to forecast the future is contained in the current parameter vector $\boldsymbol{\theta_t}$. Also, given just $\mathcal{D}_t$, all the information concerning the future is contained in the posterior distribution of $\boldsymbol{\theta_t}|\mathcal{D}_t$. Further, if this distribution is normal, $\mathcal{N}(m_t, C_t)$, then given $\mathcal{D}_t$, the pair $\{m_t, C_t\}$ only contains all the relevant information about the future.

Another crucial property of DLMs is the superposition principle: **any linear combination of independent normal DLMs is a normal DLM**. The important consequence is that in most practical cases, a DLM can be decomposed into a linear combination of simple canonical DLMs.

Definition 5. *Let Y_t be a vector of observations, the general univariate normal DLM is defined by the quadruple $\{F_t, G_t, \sigma_t^2, W_t\}$ for each time t. This quadruple defines the model between Y_t and the parameter vector $\boldsymbol{\theta_t}$ by means of the following distributions*

$$\begin{aligned} Y_t|\boldsymbol{\theta_t}, \mathcal{D}_{t-1} &\sim \mathcal{N}(F_t^T \boldsymbol{\theta_t}, \sigma_t^2) \\ \boldsymbol{\theta_t}|\boldsymbol{\theta_{t-1}}, \mathcal{D}_{t-1} &\sim \mathcal{N}(G_t \boldsymbol{\theta_{t-1}}, W_t) \end{aligned}$$

Equivalently

$$\begin{aligned} Y_t &= F_t^T \boldsymbol{\theta_t} + v_t & v_t &\sim \mathcal{N}(0, \sigma_t^2) & &\textit{Observation equation} \\ \boldsymbol{\theta_t} &= G_t \boldsymbol{\theta_{t-1}} + \boldsymbol{\omega_t} & \boldsymbol{\omega_t} &\sim \mathcal{N}(\mathbf{0}, W_t) & &\textit{System equation} \end{aligned}$$

We assume the error sequences v_t and $\boldsymbol{\omega_t}$ to be internally and mutually independent.

This equations relates the response Y_t to θ_t via a dynamic linear regression with a normal error structure having known, though possibly time varying, observational variance σ_t^2. For time t:

- F_t is an $n \times 1$ design matrix of known values of independent variables
- θ_t is a $n \times 1$ state vector
- G_t is an $n \times n$ known matrix, describing the state dynamic, named evolution matrix
- W_t is an $n \times n$ known variance matrix

Note that, given θ_t, the distribution of the response Y_t is independent of any past observation $Y_{t-1}, Y_{t-2}, \ldots, Y_0$ and $\theta_{t-1}, \theta_{t-2}, \ldots, \theta_0$. In the same way, given θ_{t-1} the distribution of the state vector is independent of any past observation $Y_{t-1}, Y_{t-2}, \ldots, Y_0$ and $\theta_{t-2}, \theta_{t-3}, \ldots, \theta_0$.

We call $\mu_t = F_t^t \boldsymbol{\theta_t}$ mean response, or level, of the series. Morover we assume that the information set at time t, namely $\mathcal{D}_t$, contains the values of σ_t^2 (if assumed known), W_t, F_t, G_t for any time $t > 0$ as well as all past response observations $Y_{t-1}, Y_{t-2}, \ldots, Y_0$.

Definition 6. *If the couple $\{F_t, G_t\}$ is constant for all t, then the DLM is referred to as **time series DLM**, or **TSDLM**. If morover the observational variance σ_t^2 and the state evolution variance W_t are constant for all t as well, the TSDLM is referred to as **constant DLM**.*

A constant DLM is characterized by the single quadruple $\{F, G, \sigma^2, W\}$.

The main result about normal DLM theory follows: intuitively the central characteristic of the normal model is that at any time, existing information about the system is represented and sufficiently summarised by the posterior distribution for the current state vector.

In the following we assume that $\mathcal{D}_0$ contains our prior knowledge about the state vector as well as the values of $\{F_t, G_t, \sigma_t^2, W_t\}$ for any t, so that at any time instant $t > 0$, the information set $\mathcal{D}_t$ is given by $\mathcal{D}_{t-1} \cup \{Y_t\}$.

Theorem 5. ***(Updating equations for the general univariate DLM)***: *Given the general univariate dynamic linear model*

$$\begin{cases} Y_t = F_t^T \boldsymbol{\theta_t} + v_t \quad v_t \sim \mathcal{N}(0, \sigma_t^2) & \textit{Observation equation} \\ \boldsymbol{\theta_t} = G_t \boldsymbol{\theta_{t-1}} + \boldsymbol{\omega_t} \quad \boldsymbol{\omega_t} \sim \mathcal{N}(\mathbf{0}, W_t) & \textit{System equation} \\ \boldsymbol{\theta_0} | \mathcal{D}_0 \sim \mathcal{N}(\boldsymbol{m_0}, C_0) & \textit{Initial information} \end{cases}$$

for any time $t > 0$ the one step forecast and posterior distributions can be obtained sequentially as:

$$\begin{array}{llll} \boldsymbol{\theta_t} | \mathcal{D}_{t-1} \sim \mathcal{N}(\boldsymbol{a_t}, R_t) & & R_t = G_t C_{t-1} G_t^T + W_t & \textit{Prior for } \boldsymbol{\theta_t} \\ & & \boldsymbol{a_t} = G_t \boldsymbol{m_{t-1}} & \\ Y_t | \mathcal{D}_{t-1} \sim \mathcal{N}(f_t, Q_t) & & Q_t = F_t^T R_t F_t + \sigma_t^2 & \textit{One step forecast} \\ & & f_t = F_t^T \boldsymbol{a_t} & \\ \boldsymbol{\theta_t} | \mathcal{D}_t \sim \mathcal{N}(\boldsymbol{m_t}, C_t) & m_t = \boldsymbol{m_{t-1}} + A_t e_t & C_t = R_t - A_t Q_t A_t^T & \textit{Posterior for } \boldsymbol{\theta_t} \\ & A_t = R_t F_t Q_t^{-1} & e_t = Y_t - f_t & \end{array}$$

Proof: By induction, let assume $\theta_{t-1} | \mathcal{D}_{t-1} \sim \mathcal{N}(\boldsymbol{m_{t-1}}, C_{t-1})$. This is true for $t = 1$ by initial condition assumptions. Using gaussian properties:

$$\begin{cases} \boldsymbol{\theta_{t-1}} | \mathcal{D}_{t-1} \sim \mathcal{N}(\boldsymbol{m_{t-1}}, C_{t-1}) \\ \omega_t \sim \mathcal{N}(\mathbf{0}, W_t) \end{cases} \qquad \boldsymbol{\theta_t} = G_t \boldsymbol{\theta_{t-1}} + \boldsymbol{\omega_t} \qquad \rightarrow \qquad \boldsymbol{\theta_t} | \mathcal{D}_{t-1} \sim \mathcal{N}(\boldsymbol{a_t}, R_t)$$

in the same way, from observation equation

$$\begin{cases} \boldsymbol{\theta_t} | \mathcal{D}_{t-1} \sim \mathcal{N}(\boldsymbol{a_t}, R_t) \\ v_t \sim \mathcal{N}(0, \sigma_t^2) \end{cases} \qquad Y_t = F_t^T \boldsymbol{\theta_t} + v_t \qquad \rightarrow \qquad Y_t | \mathcal{D}_{t-1} \sim \mathcal{N}(f_t, Q_t)$$

To derive the posterior distribution of $\boldsymbol{\theta_t}$, note that

$$\begin{aligned} COV[\boldsymbol{\theta_t}, Y_t | \mathcal{D}_{t-1}] &= COV[\boldsymbol{\theta_t}, F_t^t \boldsymbol{\theta_t} + v_t | \mathcal{D}_{t-1}] = COV[\boldsymbol{\theta_t}, F_t^t \boldsymbol{\theta_t} | \mathcal{D}_{t-1}] + COV[\boldsymbol{\theta_t}, v_t | \mathcal{D}_{t-1}] \overset{\boldsymbol{\theta_t} \perp v_t}{=} \\ &= COV[\boldsymbol{\theta_t}, \boldsymbol{\theta_t} | \mathcal{D}_{t-1}] F_t = VAR[\boldsymbol{\theta_t} | \mathcal{D}_{t-1}] F_t = R_t F_t \end{aligned}$$

then let us consider the joint distribution of $\boldsymbol{\theta_t}$ and Y_t conditioned on $\mathcal{D}_{t-1}$. Note that, given α and β in $\mathbb{R}$:

$$\boldsymbol{\alpha^T} \boldsymbol{\theta_t} + \beta Y_t = \boldsymbol{\alpha^T} \boldsymbol{\theta_t} + \beta (F_t^T \boldsymbol{\theta_t} + v_t) = (\boldsymbol{\alpha^T} + \beta F_t^T) \boldsymbol{\theta_t} + \beta v_t$$

which has gaussian distribution once conditioned on $\mathcal{D}_{t-1}$ since sum of two independent gaussian random variables (note that $\boldsymbol{\theta_t} | \mathcal{D}_{t-1}$ is jointly gaussian, hence is guaranteed that $(\boldsymbol{\alpha^T} + \beta F_t^T) \boldsymbol{\theta_t}$ is gaussian too, since linear combination of elements of a gaussian random vector). Hence because any linear combination of $\boldsymbol{\theta_t}$ and Y_t has gaussian distribution, their joint distribution is gaussian. So that we can write

$$\begin{bmatrix} \boldsymbol{\theta_t} \\ Y_t \end{bmatrix} | \mathcal{D}_{t-1} \sim \mathcal{N}_{n+1} \left(\begin{bmatrix} \boldsymbol{a_t} \\ f_t \end{bmatrix}, \begin{bmatrix} R_t & R_t F_t \\ F_t^T R_t & Q_t \end{bmatrix} \right)$$

hence the conditional distribution of $\boldsymbol{\theta_t}$ given Y_t is

$$\boldsymbol{\theta_t} | Y_t, \mathcal{D}_{t-1} = \boldsymbol{\theta_t} | \mathcal{D}_t \sim \mathcal{N}(\boldsymbol{a_t} + R_t F_t Q_t^{-1} (Y_t - f_t), R_t - (R_t F_t Q_t^{-1}) Q_t (R_t F_t Q_t^{-1})^{-1})$$

Once set $A_t = R_t F_t Q_t^{-1}$ and $e_t = Y_t - f_t$ the result follow.

So far, the defining quadruples of the univariate DLM have been assumed known for all time. Generally, the regression vectors F_t and the evolution matrices G_t are defined as part of the modeling phase. The evolution variance matrix is also chosen during this phase, usually applying the **discount principle** (???????????). However the remaining element of each quadruple, the observational variance σ_t^2, is often unknown. We present now the model for the case where σ_t^2 is unknown. We assume as always a conjugate prior and we focus on the case of constant observational variance $\sigma_t^2 = \sigma^2$

Definition 7. *For each t, the general univariate DLM in case of unknown but constant observational variance σ^2 is defined by*

$$\begin{cases} Y_t = F_t^T \boldsymbol{\theta_t} + v_t \quad v_t \sim \mathcal{N}(0, \sigma^2) & \textit{Observation equation} \\ \boldsymbol{\theta_t} = G_t \boldsymbol{\theta_{t-1}} + \boldsymbol{\omega_t} \quad \boldsymbol{\omega_t} \sim \mathcal{N}(\mathbf{0}, \sigma^2 W_t) & \textit{System equation} \\ \boldsymbol{\theta_0} | \mathcal{D}_0, \sigma^2 \sim \mathcal{N}(\boldsymbol{m_0}, \sigma^2 C_0) & \textit{Initial information} \\ \sigma^2 | \mathcal{D}_0 \sim inv\Gamma\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right) & \end{cases}$$

for some known m_0, C_0, n_0 and S_0. Matrices $\{F_t, G_t W_t\}$ are also assumed known. The error sequences v_t and ω_t are internally independent and mutually independent, and independent of intial information $\mu_0|\mathcal{D}_0$, once conditioned on σ^2.

In the above definition S_0 is a prior point estimate for the observational variance σ^2. We report the update equations for this case

Theorem 6. ***(Updating equations for the general univariate DLM with unknown observational variance)****: Given the general univariate dynamic linear model*

$$\begin{cases} Y_t = F_t^T \boldsymbol{\theta_t} + v_t \quad v_t \sim \mathcal{N}(0, \sigma^2) & \textit{Observation equation} \\ \boldsymbol{\theta_t} = G_t \boldsymbol{\theta_{t-1}} + \boldsymbol{\omega_t} \quad \boldsymbol{\omega_t} \sim \mathcal{N}(\mathbf{0}, \sigma^2 W_t) & \textit{System equation} \\ \boldsymbol{\theta_0} | \mathcal{D}_0, \sigma^2 \sim \mathcal{N}(\boldsymbol{m_0}, \sigma^2 C_0) & \textit{Initial information} \\ \sigma^2 | \mathcal{D}_0 \sim inv\Gamma\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right) & \end{cases}$$

for any time $t > 0$ the one step forecast and posterior distributions can be obtained sequentially as:

$$\begin{aligned}
&\sigma^2 | \mathcal{D}_{t-1} \sim inv\Gamma\left(\frac{n_{t-1}}{2}, \frac{n_{t-1} S_{t-1}}{2}\right) && && \textit{Prior for } \sigma^2 \\
&\boldsymbol{\theta_t} | \mathcal{D}_{t-1}, \sigma^2 \sim \mathcal{N}(\boldsymbol{a_t}, \sigma^2 R_t) && R_t = G_t C_{t-1} G_t^T + W_t && \textit{Prior for } \boldsymbol{\theta_t} \\
& && \boldsymbol{a_t} = G_t \boldsymbol{m_{t-1}} && \\
&Y_t | \mathcal{D}_{t-1}, \sigma^2 \sim \mathcal{N}(f_t, \sigma^2 Q_t) && Q_t = F_t^T R_t F_t + 1 && \textit{One step forecast} \\
& && f_t = F_t^T \boldsymbol{a_t} && \\
&\sigma^2 | \mathcal{D}_t \sim inv\Gamma\left(\frac{n_t}{2}, \frac{n_t S_t}{2}\right) && n_t = n_{t-1} + 1 && \textit{Posterior for } \sigma^2 \\
& && n_t S_t = n_{t-1} S_{t-1} + e_t^2 Q_t^{-1} && \\
&\boldsymbol{\theta_t} | \mathcal{D}_t, \sigma^2 \sim \mathcal{N}(\boldsymbol{m_t}, \sigma^2 C_t) && m_t = \boldsymbol{m_{t-1}} + A_t e_t \quad C_t = R_t - A_t Q_t A_t^T && \textit{Posterior for } \boldsymbol{\theta_t} \\
& && A_t = R_t F_t Q_t^{-1} \quad e_t = Y_t - f_t &&
\end{aligned}$$

Principle of superposition

We pose the problem of defining system matrices F_t and G_t. In applications, models are usually constructed by combining two or more component DLMs, each of which captures an individual feature of the real series under study. The construction of complex DLMs from component DLMs is referred to as superposition.

Theorem 7. ***(Superposition)****: Consider h time series $Y_{it} \quad i = 1, \dots, h$, each one generated by a DLM $\mathcal{M}_i$ described by the quadruple $\{F_{it}, G_{it}, \sigma_{it}^2, W_{it}\}$. Let assume the state vector $\boldsymbol{\theta_{it}}$ of model $\mathcal{M}_i$ to be of dimension $n_i \times 1$. Denote with v_{it} and $\boldsymbol{\omega_{it}}$ the observational and evolution error od model $\mathcal{M}_i$. The state vectors are distinct, and for all distinct $i \neq j$, the series v_{it} and ω_{it} are mutually independent of the series v_{jt} and ω_{jt}. Then the series*

$$Y_t = \sum_{i=1}^{h} Y_{it}$$

follows the n-dimensional DLM $\{F_t, G_t, \sigma_t^2, W_t\}$ where $n = \sum_{i=1}^{h} n_i$ and

$$F_t = \begin{bmatrix} F_{1t} \\ F_{2t} \\ \vdots \\ F_{ht} \end{bmatrix} \qquad G_t = \begin{bmatrix} G_{1t} & 0 & \dots & 0 \\ 0 & G_{2t} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & G_{ht} \end{bmatrix} \qquad \sigma_t^2 = \sum_{i=1}^{h} \sigma_{it}^2 \qquad W_t = \begin{bmatrix} W_{1t} & 0 & \dots & 0 \\ 0 & W_{2t} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & W_{ht} \end{bmatrix}$$

and the state vector $\boldsymbol{\theta_t}$ is obtained by concatenating the state vectors of each single model $\mathcal{M}_i$.

The final message is: **a linear combination of DLMs is a DLM**. In practice, any complex target series Y_t we want to forecast can be seen as linear combination of simple components: a **trend** describing the slow variation dynamic of the series, a **seasonal** component, capturing periodical patterns inside the data, a **regressive** and **autoregressive** component.

Polynomial trend models

In time series these models prove useful in describing trends that are generally viewed as smooth developments over time. Relative to the sampling interval of the series and the required forecast horizons, such trends are usually well approximated by low-order polynomial functions of time. Indeed, a first or second order polynomial component DLM is often quite adequate for short term forecasting.

Polynomial DLMs are a subset of TSDLM, hence both F_t and G_t are constant over time and equal to F and G respectively.

Definition 8. *Any TSDLM which for all times $t > 0$ has a forecast function of the form*

$$f_t(k) = \mathbb{E}[Y_{t+k}|\mathcal{D}_t] = a_{t,0} + a_{t,1}k + a_{t,2}k^2 + \ldots + a_{t,n-1}k^{n-1}$$

is named ***n-order polynomial DLM***

In practice we are saying that **locally in time** the expected value of the time series evolves

- linearly as $a_{t,0} + a_{t,1}k$ if we consider a **second order polynomial model**
- quadratically as $a_{t,0} + a_{t,1}k + a_{t,2}k^2$ if we consider a **third order polynomial model**
- ...

Anyway the approximation should be taken as good only for **short term forecasts**. Coefficients $a_{t,0}, a_{t,1}, \ldots, a_{t,n-1}$ changes as new information is collected so to track the series (as result of the recursive update equations of DLMs).

A DLM is defined once system matrices F, G and w_t are defined. It can be prooven that a DLM with the following system matrices

$$F = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \end{bmatrix} \qquad G = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 0 & 1 & 1 & \dots & 1 \\ 0 & 0 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots 1 & \end{bmatrix} \qquad W_t = \begin{bmatrix} w_{t,1} & 0 & 0 & \dots & 0 \\ 0 & w_{t,2} & 0 & \dots & 0 \\ 0 & 0 & w_{t,3} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & 0 & \dots & w_{t,n} \end{bmatrix}$$

defines a DLM with a forecast function equal to the one in definition 8. The model equations are as follow

$$\begin{cases} Y_t = \theta_{t,1} + v_t \qquad v_t \sim \mathcal{N}(0, \sigma^2) & \text{Observation equation} \\ \theta_{t,j} = \theta_{t-1,j} + \displaystyle\sum_{r=j+1}^{n} \theta_{t-1,r} + \omega_{t,j} \qquad \omega_{t,j} \sim \mathcal{N}(0, w_{t,j}) \qquad j = 1, \ldots, n-1 & \text{System equations} \\ \theta_{t,n} = \theta_{t-1,n} + \omega_{t,n} \qquad \omega_{t,n} \sim \mathcal{N}(0, w_{t,n}) & \end{cases}$$

About the meaning of this model, state parameters $\theta_{t,1}, \ldots, \theta_{t,n}$ can be seen as the derivatives of the level $\theta_{t,1}$, in the sense that for each $j = 1, \ldots, n$ the term $\theta_{t,j}$ represent the j-th derivative of the series level. The next important example shows the reason for this interpretation.

Example 3. *Second-order polynomial trend model*

At any time t, a second-order polynomial DLM has a straight line forecast function of the form

$$f_t(k) = a_{t,0} + a_{t,1}k$$

According to the general formulation of an n-th order polynomial model we have:

$$F = \begin{bmatrix} 1 & 0 \end{bmatrix} \qquad G = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \qquad W_t = \begin{bmatrix} w_{t,1} & 0 \\ 0 & w_{t,2} \end{bmatrix}$$

Let consider the following parametrization

$$\boldsymbol{\theta_t} = \begin{bmatrix} \theta_{t,1} \\ \theta_{t,2} \end{bmatrix} = \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix}$$

where μ_t is the level of the series while β_t takes the meaning of increment of the level μ_t between time $t-1$ and t (indeed, the first derivative of the level).

Definition 9. *For each t, the second order polynomial DLM is defined by*

$$\begin{cases} Y_t = \mu_t + v_t \quad v_t \sim \mathcal{N}(0, \sigma_t^2) & \textit{Observation equation} \\ \mu_t = \mu_{t-1} + \beta_{t-1} + \omega_{t,1} \quad \omega_{t,1} \sim \mathcal{N}(0, w_{t,1}) & \textit{System equation} \\ \beta_t = \beta_{t-1} + \omega_{t,2} \quad \omega_{t,2} \sim \mathcal{N}(0, w_{t,2}) & \\ \boldsymbol{\theta_0} | \mathcal{D}_0 \sim \mathcal{N}(m_0, C_0) & \textit{Initial information} \end{cases}$$

where the error sequences v_t and $\boldsymbol{\omega_t}$ are internally independent and mutually independent, and independent of intial information $\boldsymbol{\theta_0}|\mathcal{D}_0$.

We can see why β_t takes the meaning of first derivative of the level μ_t. Indeed, let us not consider the stochastic variation affecting μ_t, thus from the system equation we can write

$$\mu_t - \mu_{t-1} = \frac{\Delta \mu_t}{\Delta t} \overset{\Delta t = 1}{=} \beta_{t-1}$$

hence β_{t-1} indicates the variation of the series level μ_t between the time instant $t-1$ and t, excluding a small fluctuation. Notice that β_t itself exhibit a slow variation dinamic, namely a random walk dynamic. Thus in such model, β_t and as a consequence the increment of μ_t, changes only by the effect of random fluctuations.

Inference is performed using general DLM theory seen in previous sections.

Seasonal models

The term seasonality is used as a label for any cyclical or periodic behaviour. Let $g(\cdot)$ be any real-valued function defined on the non-negative integers $t = 1, 2, \ldots$, where t is a time index. $g(t)$ is periodic if there exists integer $p > 1$ such that $g(t + np) = g(t)$ for all integers t and $n \geq 0$. We call p **period** of $g(\cdot)$.

The **seasonal factors** of $g(\cdot)$ are the p values taken in any time interval containing p consecutive time points, such as $[t, t+p-1]$, for any $t > 0$. We denote the seasonal factors as

$$\psi_j = g(j) \qquad j = 1, \ldots, p-1$$

Thanks to periodicity, for $t > 0$ we have $g(t) = g(j)$ where j is the remainder of the division of t by p, denoted by $j = t|p$.

We call **seasonal factor vector** at time t the permutation of the vector of seasonal factors that has its first element relating to time t. You can imagine the seasonal factor vector to rotate as time passes:

$$\begin{aligned} \boldsymbol{\psi_0} &= [\psi_0, \psi_1, \ldots, \psi_{p-1}]^T & j &= 0|p \\ &\vdots \\ \boldsymbol{\psi_j} &= [\psi_j, \ldots, \psi_0, \ldots, \psi_{j-1}]^T & j &= j|p \\ &\vdots \\ \boldsymbol{\psi_p} &= [\psi_0, \psi_1, \ldots, \psi_{p-1}]^T & j &= p|p \\ &\vdots \end{aligned}$$

When the p seasonal factors relating to a period p may take arbitrary real values, the seasonal pattern is termed **form-free**. Hence the idea is clear: at any time t, if the effect on the output given by the seasonal component is only due by the element in first position of the seasonal factor vector, to build a seasonal DLM is just a matter of rotating the elements of this vector, which will constitute the state of the system.

Definition 10. *We define $\boldsymbol{E_p}$ and the $p \times p$ permutation matrix P as*

$$\boldsymbol{E_p} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \qquad P = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 0 \end{bmatrix}$$

It can be prooven that the matrix P so defined is p-cyclic, namely $\forall n \geq 0, P^{np} = I_p$ and $P^{k+np} = P^k, k = 1, \dots, p$.

Using $\boldsymbol{E_p}$ and the permutation matrix p, we have that at any time t

$$\psi_j = g(t) = \boldsymbol{E_p^T \psi_t} \qquad j = t|p$$

while the seasonal factors rotate according to

$$\boldsymbol{\psi_t} = P\boldsymbol{\psi_{t-1}}$$

This relationship provides the initial step in constructing a purely seasonal DLM

Definition 11. ***(Form-free seasonal factor DLM)****: the canonical form-free seasonal factor DLM of period $p > 1$ is defined, for any appropriate variances σ_t^2 and W_t, by the quadruple $\{\boldsymbol{E_p}, P, \sigma_t^2, W_t\}$.*

Letting the state vector be the seasonal factor vector $\boldsymbol{\psi_t}$, the DLM equations are

$$\begin{cases} Y_t = \boldsymbol{E_p^T \psi_t} + v_t \quad v_t \sim \mathcal{N}(0, \sigma_t^2) & \textit{Observation equation} \\ \boldsymbol{\psi_t} = P\boldsymbol{\psi_{t-1}} + \boldsymbol{\omega_t} \quad \boldsymbol{\omega_t} \sim \mathcal{N}(\boldsymbol{0}, W_t) & \textit{System equation} \end{cases}$$

The previous definition is usefull in describing a purely seasonal model. Generally we decompose a set of p seasonal factors in two components: a deseasonalised level of the series and p seasonal deviations from this level. The seasonal deviations from the series level are called **seasonal effects**.

This may be thought as a classical model for treatment effect (LMM). We have a common mean for all observations (the series level) and the effect of each treatment (the seasonal effects) will act on the common mean when a group condition is met. Due to identifiability problems, we need to constraint the treatment effect vector. The commonest contraint is the zero-sum constraint. The analogy for seasonality is that the seasonal deviations from the underlying level sum to zero in a full period.

Initially, the underlying level of the series is set to zero for all t, so that the seasonal factors always sum to zero, producing a seasonal effects DLM. The superposition of this seasonal effect DLM and a first-order polynomial DLM then provides the constrained, seasonal effects component for a series with non-zero level.

Definition 12. ***(Form-free seasonal effects DLM)****: A form free seasonal effects DLM is any model described by the quadruple $\{\boldsymbol{E_p}, P, \sigma_t^2, W_t\}$ with state vector $\phi_t = [\phi_{t,0}, \dots, \phi_{t,p-1}]$ satisfying $\sum_{i=0}^{p-1} \phi_{t,i} = 0$*

The seasonal effects $\phi_{t,j}$ represent seasonal deviations from their zero mean and are simply constrained seasonal factors.

Trend form-free seasonal effects DLM

The main reason for considering the form-free seasonal effects DLM is that it provides a widely applicable seasonal component that in a larger DLM, describes seasonal deviations from a deseasonalised level, or trend.

Example 4. *First-order polynomial seasonal effects model*

A **first-order polynomial trend form-free seasonal effects DLM** is any DLM with parameter vector

$$\boldsymbol{\theta_t} = \begin{bmatrix} \mu_t \\ \boldsymbol{\phi_t} \end{bmatrix}$$

and quadruple

$$\left\{ \begin{bmatrix} 1 \\ \boldsymbol{E_p} \end{bmatrix}, \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & P \end{bmatrix}, \sigma_t^2, \begin{bmatrix} W_{t,\mu} & \mathbf{0} \\ \mathbf{0} & W_{t,\phi} \end{bmatrix} \right\}$$

satisfying $\sum_{i=0}^{p-1} \phi_{t,i} = 0$ for all t. Such models comprise the superposition of a first-order polynomial DLM (for the deseasonalised level) and a seasonal effects DLM. The forecast function takes the form

$$f_t(k) = m_t + h_{tj} \qquad j = k|p$$

where m_t is the expected value of the deseasonalised level at time $t+k$ and $h_{t,j}$ is the expected seasonal deviation from this level. The model can be also explicitly written as

$$\begin{cases} Y_t = \mu_t + \phi_{t,0} + v_t \quad v_t \sim \mathcal{N}(0, \sigma_t^2) & \text{Observation equation} \\ \mu_t = \mu_{t-1} + \omega_t \quad \omega_t \sim \mathcal{N}(0, w_t) & \text{Trend dynamic} \\ \phi_{t,r} = \phi_{t-1,r+1} + \omega_{t,r} \quad \omega_{t,r} \sim \mathcal{N}(0, w_{t,r}) \quad r = 1, \ldots, p-2 & \text{Seasonal dynamic} \\ \phi_{t,p-1} = \phi_{t-1,0} + \omega_{t,p-1} \quad \omega_{t,p-1} \sim \mathcal{N}(0, w_{t,p-1}) & \end{cases}$$

with the zero sum constraint on the seasonal effects.

Example 5. *Second-order polynomial seasonal effects model*

A first-order polynomial trend form-free seasonal effects DLM is any DLM with parameter vector

$$\boldsymbol{\theta_t} = \begin{bmatrix} \mu_t \\ \beta_t \\ \boldsymbol{\phi_t} \end{bmatrix}$$

and quadruple

$$\left\{ \begin{bmatrix} \boldsymbol{E_2} \\ \boldsymbol{E_p} \end{bmatrix}, \begin{bmatrix} 1 & 1 & \mathbf{0} & \\ 0 & 1 & \mathbf{0}, \mathbf{0} & P \end{bmatrix}, \sigma_t^2, \begin{bmatrix} W_{t,\mu} & \mathbf{0} \\ \mathbf{0} & W_{t,\phi} \end{bmatrix} \right\}$$

satisfying $\sum_{i=0}^{p-1} \phi_{t,i} = 0$ for all t. Such models comprise the superposition of a second-order polynomial DLM (for the deseasonalised level) and a seasonal effects DLM. The forecast function takes the form

$$f_t(k) = m_t + kb_t + h_{tj} \qquad j = k|p$$