Chapter 5

Classical time series

In this chapter, we will consider the different models that are widespread in the time series literature, that is the autoregressive (AR), moving-average (MA) models as well as the standard combinations of such models such as the ARMA model and the ARIMA model where the "I" stands for integrated. However, we will start with an important notion in time-series analysis which will be relevant for the considered models: stationarity.

5.1 Stationarity

The assumption of stationarity is important in time series analysis since it guarantees that the mean and variance do not change over time. We consider a time series $(\boldsymbol{y}_k)_{k\geq 0}$ characterised by a given model. For $(\boldsymbol{y}_k)_{k\geq 0}$ to be stationary, the joint distribution of this time series has to verify

$$p_{\boldsymbol{y}_{k_1},\ldots,\boldsymbol{y}_{k_n}}(\cdot) = p_{\boldsymbol{y}_{k_1+\delta},\ldots,\boldsymbol{y}_{k_n+\delta}}(\cdot)$$

that is the joint probability density function characterising the time series at some time steps $k_1, k_2, \ldots, k_n$ for some $n > 0$ has to be invariant under a time shift of $\delta \geq 0$ steps. This condition is strong and might be difficult to verify in practice so we also introduce a relaxed version called *weak stationarity*, which simply assumes that $\mathbb{E}(\boldsymbol{y}_k)$ and $\mathrm{var}(\boldsymbol{y}_k)$ are constant with respect to k and that

$$\mathrm{cov}(\boldsymbol{y}_k, \boldsymbol{y}_{k-\delta}) = \mathbb{E}\big([\boldsymbol{y}_k - \mathbb{E}(\boldsymbol{y}_k)][\boldsymbol{y}_{k-\delta} - \mathbb{E}(\boldsymbol{y}_{k-\delta})]\big) = \gamma_\delta$$

for any $k \geq 0$ and any $\delta < k$, where γ_δ is referred to as the *auto-covariance*. Note that if the second moment of a time series exists then stationarity implies weak stationarity, that is, stationarity is a stronger assumption than weak stationarity, as the name suggest. Yet, in the case of a Gaussian DLM, the two notions are equivalent (this is not surprising since normal distributions are characterised by their first two moments).

Since the variance of the (weakly) stationary time series $(\boldsymbol{y}_k)_{k\geq 0}$ is constant in time, the *autocorrelation* ρ_δ can be easily deduced from the auto-covariance as

$$\rho_\delta = \frac{\mathbb{E}\big([\boldsymbol{y}_{k-\delta} - \mathbb{E}(\boldsymbol{y}_{k-\delta})][\boldsymbol{y}_k - \mathbb{E}(\boldsymbol{y}_k)]\big)}{\sqrt{\mathrm{var}(\boldsymbol{y}_{k-\delta})}\sqrt{\mathrm{var}(\boldsymbol{y}_k)}} = \frac{\gamma_\delta}{\mathrm{var}(\boldsymbol{y}_0)}$$

which is indeed independent of the time step k.

5.2 Stationary time series

The importance of the concept of weak stationarity can also be seen in the following decomposition result, where we consider time series on the set of integers $\mathbb{Z}$, that is including negative time steps.

Theorem 5.1 (Wold representation theorem). *Any weakly stationary time series $(\boldsymbol{y}_k)_{k\in\mathbb{Z}}$ can be written as*

$$\boldsymbol{y}_k = \mu + \sum_{\delta\geq 0} \psi_\delta \boldsymbol{\epsilon}_{k-\delta}$$

for some scalar μ, some coefficients $\psi_0, \psi_1, \ldots$ with $\psi_0 = 1$ and some sequence $(\boldsymbol{\epsilon}_k)_{k\in\mathbb{Z}}$ of zero-mean, uncorrelated random variables with constant variance.

Since a weakly stationary time series $(\boldsymbol{y}_k)_{k\in\mathbb{Z}}$ has constant variance, it must hold that

$$\mathrm{var}(\boldsymbol{y}_k) = \mathrm{var}(\boldsymbol{\epsilon}_0) \sum_{\delta\geq 0} \psi_\delta^2 = \mathrm{cst}$$

which implies that the sequence $(\psi_i)_{i\geq 0}$ is square-summable (formally in ℓ^2). The decomposition of the time series $(\boldsymbol{y}_k)_{k\in\mathbb{Z}}$ can also be expressed

$$\boldsymbol{y}_k = \mu + \sum_{\delta\geq 0} \psi_\delta B^\delta \boldsymbol{\epsilon}_k$$

where B is the *backshift* operator such that[1] $B\boldsymbol{y}_k = \boldsymbol{y}_{k-1}$ and $B^\delta \boldsymbol{y}_k = \boldsymbol{y}_{k-\delta}$ for any $\delta > 0$. This expression can be written more compactly as $\boldsymbol{y}_k = \mu + \psi(B)\boldsymbol{\epsilon}_k$ by introducing the polynomial function

$$\psi(x) = 1 + \psi_1 x + \psi_2 x^2 + \ldots$$

for any real x such that $|x| < 1$. Finally, if there exists a polynomial function $\phi(\cdot)$ of the form $\phi(x) = 1 - \phi_1 x - \phi_2 x^2 - \ldots$ such that $\psi(x)\phi(x) = 1$ for any x such that $|x| < 1$ then ψ is said to be *invertible* and the expression of the time series $(\boldsymbol{y}_k)_{k\in\mathbb{Z}}$ can be changed once more to $\phi(B)(\boldsymbol{y}_k - \mu) = \boldsymbol{\epsilon}_k$.

5.2.1 The autoregressive and moving-average models

If the sequence of coefficients $\phi_1, \phi_2, \ldots$ is such that $\phi_\delta = 0$ for any δ strictly greater than an integer $p > 0$ then the time series $(\boldsymbol{y}_k)_{k\in\mathbb{Z}}$ is said to be *autoregressive process of order* p, denoted AR(p), and expressed as

$$\boldsymbol{y}_k = \mu + \sum_{\delta=1}^{p} \phi_\delta(\boldsymbol{y}_{k-\delta} - \mu) + \boldsymbol{\epsilon}_k.$$

The case where $p = 1$ provides useful insights on the behaviour of autoregressive processes. Indeed, in this case:

- The expression of the process simplifies to $\boldsymbol{y}_k = \mu + \phi_1(\boldsymbol{y}_{k-1} - \mu) + \boldsymbol{\epsilon}_k$
- The variance is such that $\mathrm{var}(\boldsymbol{y}_0) = \mathrm{var}(\boldsymbol{\epsilon}_0)/(1 - \phi_1^2)$ given that $|\phi_1| < 1$
- The autocorrelation is $\rho_\delta = \phi_1^\delta$
- The corresponding polynomial function $\psi(\cdot)$ is

  $$\psi(x) = 1/\phi(x) = 1 + \phi_1 x + \phi_1^2 x^2 + \ldots$$

  for any real x such that $|x| < 1$
- The random variable $\boldsymbol{y}_k$ can be expressed as a function of $\boldsymbol{y}_0$ and the noise sequence $(\boldsymbol{\epsilon}_k)_{k\geq 0}$ as

  $$\begin{aligned}\boldsymbol{y}_k &= \mu + \phi_1(\boldsymbol{y}_{k-1} - \mu) + \boldsymbol{\epsilon}_k \\ &= \mu + \phi_1(\phi_1(\boldsymbol{y}_{k-2} - \mu) + \boldsymbol{\epsilon}_{k-1}) + \boldsymbol{\epsilon}_k \\ &= [\ldots] \\ &= \mu + \phi_1^k(\boldsymbol{y}_0 - \mu) + \sum_{i=0}^{k-1} \phi_1^i \boldsymbol{\epsilon}_{k-i} \\ &= (1 - \phi_1^k)\mu + \phi_1^k \boldsymbol{y}_0 + \sum_{i=0}^{k-1} \phi_1^i \boldsymbol{\epsilon}_{k-i}\end{aligned}$$

  from which it can be deduced that $|\phi_1| < 1$ yields a stationary process (shocks will die out with time), whereas $|\phi_1| = 1$ gives a random walk (shocks are permanent) and $|\phi_1| > 1$ makes the process unstable (shocks are amplified over time even if $\mu = 0$).

[1] the backshift operator is formally defined as $B(\boldsymbol{y}_\cdot)(k) = \boldsymbol{y}_{k-1}$ where $\boldsymbol{y}_\cdot$ is the function on $\mathbb{Z}$ defined as $\boldsymbol{y}_\cdot : k \mapsto \boldsymbol{y}_k$. The term B^δ is understood as the δ-fold composition of B.

It follows that the variance gets larger as ϕ_1 gets closer to 1 and the autocorrelation ρ_δ decays exponentially in δ.

Considering again the expression of $\boldsymbol{y}_k$ based on the coefficients $\psi_0, \psi_1, \dots,$ we can take a similar approach and assume that $\psi_\delta = 0$ for any δ strictly greater than an integer $q > 0$. In this case, the time series $(\boldsymbol{y}_k)_{k\in\mathbb{Z}}$ is said to be *moving-average process of order* q, denoted MA(q), and expressed as

$$\boldsymbol{y}_k = \mu + \sum_{\delta=0}^{q} \psi_\delta \boldsymbol{\epsilon}_{k-\delta}$$

with $\psi_0 = 1$ as before. Since $\boldsymbol{y}_k$ and $\boldsymbol{y}_{k-\delta}$ depend on different components of the sequence $(\boldsymbol{\epsilon}_k)_{k\in\mathbb{Z}}$ if $\delta > q$, it is easy to verify that $\rho_\delta = 0$ for any such δ. This is a crucial difference between AR and MA processes (remember that even AR(1) processes have $\rho_\delta > 0$ for all δ). It follows that MA processes are useful for modelling local dependencies, at least for small values of q.

To illustrate the properties of MA(q) processes, we consider the case of $q = 1$:

- The expression of the process simplifies to $\boldsymbol{y}_k = \mu + \boldsymbol{\epsilon}_k + \psi_1 \boldsymbol{\epsilon}_{k-1}$
- The (constant) variance is $\mathrm{var}(\boldsymbol{y}_0) = (1 + \psi_1^2)\,\mathrm{var}(\boldsymbol{\epsilon}_0)$
- The 1-lag autocorrelation is $\rho_1 = \psi_1/(1 + \psi_1^2)$
- The corresponding polynomial function $\phi(\cdot)$ is

  $$\phi(x) = 1/\psi(x) = 1 - \psi_1 x + \psi_1^2 x^2 - \dots$$

  for any real x such that $|x| < 1$

If one wants to fit an MA(1) model based on the 1-lag autocorrelation ρ_1, an ambiguity will arise since both the coefficients ψ_1 and ψ_1^{-1} yield the same value of ρ_1. Considering the associated expression of $\boldsymbol{y}_k$ based on the polynomial function $\phi(\cdot)$, it appears that $|\psi_1| > 1$ generates unwanted large dependencies on past values of the time-series. A coefficient verifying $|\psi_1| < 1$ is therefore preferred.

5.2.2 The autoregressive moving-average model

Although both AR(p) and MA(q) models can approximate any time series by taking the parameters p and q large enough, such an approximation might also involve a large number of coefficients. Since these two models complement each other, it is often appropriate to consider them both simultaneously and the corresponding model is called an *autoregressive moving-average* model or ARMA model. Formally, a time series $(\boldsymbol{y}_k)_{k\in\mathbb{Z}}$ of the form

$$\boldsymbol{y}_k = \mu + \sum_{\delta=1}^{p} \phi_\delta(\boldsymbol{y}_{k-\delta} - \mu) + \sum_{\delta=1}^{q} \psi_\delta \boldsymbol{\epsilon}_{k-\delta} + \boldsymbol{\epsilon}_k$$

can be introduced for some $p, q \geq 0$ with either $p > 0$ or $q > 0$, which will be referred to as ARMA(p, q).

An ARMA(p, q) model, and hence an AR(p) or MA(q) model, can be expressed as a DLM by identification of a suitable parameter vector $\boldsymbol{\theta}_k$ together with the corresponding evolution and observation matrices F and H and noise terms $\boldsymbol{u}_k$ and $\boldsymbol{v}_k$. For the sake of simplicity, we consider $\mu = 0$ and introduce $d = \max\{p, q + 1\}$, so that $\boldsymbol{y}_k$ can be expressed as

$$\boldsymbol{y}_k = \sum_{\delta=1}^{d} \big(\phi_\delta \boldsymbol{y}_{k-\delta} + \psi_\delta \boldsymbol{\epsilon}_{k-\delta}\big) + \boldsymbol{\epsilon}_k$$

with $\phi_\delta = 0$ when $p < \delta \leq d$ and $\psi_\delta = 0$ when $q < \delta \leq d$. The associated DLM, of dimension d, can be expressed as

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

that is with $\boldsymbol{v}_k = 0$, where

$$F = \begin{pmatrix} \phi_1 & 1 & 0 & \dots & 0 \\ \phi_2 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_{d-1} & 0 & 0 & \dots & 1 \\ \phi_d & 0 & 0 & \dots & 0 \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{u}_k = \begin{pmatrix} 1 \\ \psi_1 \\ \psi_2 \\ \vdots \\ \psi_{d-1} \end{pmatrix} \boldsymbol{\epsilon}_k.$$

It is easy to verify that this DLM yields the right form for $\boldsymbol{y}_k$ as follows: the observation equation gives $\boldsymbol{y}_n = \boldsymbol{\theta}_{n,1}$ for any n, and it follows from the first component of the evolution equation at time step k that

$$\begin{aligned} \boldsymbol{y}_k &= \phi_1 \boldsymbol{\theta}_{k-1,1} + \boldsymbol{\theta}_{k-1,2} + \boldsymbol{\epsilon}_k \\ &= \phi_1 \boldsymbol{y}_{k-1} + \boldsymbol{\theta}_{k-1,2} + \boldsymbol{\epsilon}_k. \end{aligned}$$

Yet, $\boldsymbol{\theta}_{k-1,2}$ can be further expressed using the evolution equation at time step $k-1$, which leads to

$$\begin{aligned} \boldsymbol{y}_k &= \phi_1 \boldsymbol{y}_{k-1} + \big(\phi_2 \boldsymbol{\theta}_{k-2,1} + \boldsymbol{\theta}_{k-2,3} + \psi_1 \boldsymbol{\epsilon}_{k-1}\big) + \boldsymbol{\epsilon}_k \\ &= \phi_1 \boldsymbol{y}_{k-1} + \phi_2 \boldsymbol{y}_{k-2} + \boldsymbol{\theta}_{k-2,3} + \psi_1 \boldsymbol{\epsilon}_{k-1} + \boldsymbol{\epsilon}_k, \end{aligned}$$

where the original expression of $\boldsymbol{y}_k$ starts to appear. Iterating this procedure obviously leads to the desired result.

Example 5.1. Consider an ARMA$(2,1)$ process with $\mu = 0$, i.e. $\boldsymbol{y}_k = \phi_1 \boldsymbol{y}_{k-1} + \phi_2 \boldsymbol{y}_{k-2} + \psi_1 \boldsymbol{\epsilon}_{k-1} + \boldsymbol{\epsilon}_k$. The associated DLM is

$$\begin{aligned} \boldsymbol{\theta}_k &= \begin{pmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{pmatrix} \boldsymbol{\theta}_{k-1} + \begin{pmatrix} 1 \\ \psi_1 \end{pmatrix} \boldsymbol{\epsilon}_k \\ \boldsymbol{y}_k &= \begin{pmatrix} 1 & 0 \end{pmatrix} \boldsymbol{\theta}_k. \end{aligned}$$

We want to apply the approach of Chapter 3 to understand how this DLM behaves. For this purpose we have to compute the eigenvalues of F, for instance via the characteristic polynomial as follows

$$|F - \lambda I_2| = \lambda^2 - \lambda\phi_1 - \phi_2.$$

The discriminant is $\Delta = \phi_1^2 + 4\phi_2$ so that the roots are

$$\begin{aligned} \lambda_1 &= \frac{1}{2}\big(\phi_1 + \sqrt{\Delta}\big) \\ \lambda_2 &= \frac{1}{2}\big(\phi_1 - \sqrt{\Delta}\big) \end{aligned}$$

which might be distinct real eigenvalues if $\Delta > 0$, a single eigenvalue of multiplicity 2 if $\Delta = 0$ or a pair of conjugate complex numbers if $\Delta < 0$. The ARMA$(2,1)$ process will therefore assume different behaviours depending on ϕ_1 and ϕ_2, the model being stable if $|\lambda_1| < 1$ and $|\lambda_2 < 1|$, i.e. we want the roots of the characteristics polynomial to be inside of the unit circle (in the complex plane). The result does not depend on ψ_1 which only affects the magnitude of the noise. Focusing on the AR(2) part of this model, we could also consider the corresponding polynomial $\phi(\cdot)$ defined as

$$\phi(x) = 1 - \phi_1 x - \phi_2 x^2.$$

The polynomial ϕ verifies $\phi(1/\lambda) = |F - \lambda I_2|$ and therefore has the same discriminant. The corresponding roots are

$$\begin{aligned} x_1 &= \frac{1}{2\phi_2}\big(\phi_1 + \sqrt{\Delta}\big) = \frac{\lambda_1}{\phi_2} \\ x_2 &= \frac{1}{2\phi_2}\big(\phi_1 - \sqrt{\Delta}\big) = \frac{\lambda_2}{\phi_2} \end{aligned}$$

which will lie outside of the unit circle whenever λ_1 and λ_2 are inside of it since $\phi_2 < 1$ for a stable AR(2) process.

Note that other DLMs can be used to write classical time-series models depending on the context. In particular, in the case of an AR(p) process with unknown coefficients, one can consider $\boldsymbol{\theta}_k = (\phi_1, \dots, \phi_p)^\intercal$ with $F = I_p$ and $\boldsymbol{u}_k = 0$, so that $\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k-1}$, as well as with H_k equals to the row vector of the p previous observations $(y_{k-1}, \dots, y_{k-p})$ and with $\boldsymbol{v}_k = \boldsymbol{\epsilon}_k$.

5.3 Non-stationary time series

There are two main reasons for a time-series to fail to be stationary. Non-stationarity can be caused by the presence of i) a polynomial trend and/or ii) a seasonal trend.

These components can however be easily removed. For instance, if a time series $(\boldsymbol{y}_k)_{k\in\mathbb{Z}}$ follows a second-order polynomial model, it will obviously be non-stationary as soon as the rate of growth is non-zero; however the time series with term $\boldsymbol{y}_k - \boldsymbol{y}_{k-1}$ at time step k will be stationary. Using once again the backshift operator B, this new time series can be written as $((1-B)\boldsymbol{y}_k)_{k\in\mathbb{Z}}$. Similarly, if $(\boldsymbol{y}_k)_{k\in\mathbb{Z}}$ follows a d^{th}-order polynomial model then the time series $((1-B)^{d-1}\boldsymbol{y}_k)_{k\in\mathbb{Z}}$ will be stationary.

Seasonal trends of period p can also be removed by subtracting another term of the series that is p time steps apart, e.g. the time series with term $\boldsymbol{y}_k - \boldsymbol{y}_{k-p}$ will not display any seasonal trend if the original time series displays a seasonal trend of period p. This can also be expressed using the backshift operator B as $((1-B^p)\boldsymbol{y}_k)_{k\in\mathbb{Z}}$.

These two transformations are referred to as *differencing* and can be combined when both polynomial and seasonal trends are present, in which case the corresponding time series will be of the form

$$\big((1-B)^n(1-B^p)\boldsymbol{y}_k\big)_{k\in\mathbb{Z}} \tag{5.5}$$

for some $n > 1$ and some $p > 1$. If the time series (5.5) follows an ARMA model then $(\boldsymbol{y}_k)_{k\in\mathbb{Z}}$ is said to be an *autoregressive integrated moving-average* (ARIMA) process.

Example 5.2. In order to illustrate differencing for time series displaying a polynomial trend, we consider a third-order polynomial trend model as follows

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

with the components of $\boldsymbol{\theta}_k$ interpreted as a position $\boldsymbol{x}_k$, a velocity $\dot{\boldsymbol{x}}_k$ and a constant acceleration $\ddot{\boldsymbol{x}}$, i.e.

$$\boldsymbol{\theta}_k = \begin{pmatrix}\boldsymbol{x}_k\\ \dot{\boldsymbol{x}}_k\\ \ddot{\boldsymbol{x}}\end{pmatrix} \qquad \text{and} \qquad \boldsymbol{u}_k = \begin{pmatrix}\boldsymbol{\epsilon}_k\\ \dot{\boldsymbol{\epsilon}}_k\\ 0\end{pmatrix}.$$

The term $\dot{\boldsymbol{\epsilon}}_k$ is interpreted as the noise on the velocity and there is no noise on the acceleration (which would not be constant otherwise). The differencing that must be applied to the time series $(\boldsymbol{y}_k)_{k\in\mathbb{Z}}$ to obtain a new stationary time-series $(\boldsymbol{y}'_k)_{k\in\mathbb{Z}}$ is

$$\boldsymbol{y}'_k = (1-B)^2\boldsymbol{y}_k = \boldsymbol{y}_k - 2\boldsymbol{y}_{k-1} + \boldsymbol{y}_{k-2}.$$

We can use the state equation to find

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

as well as

$$\boldsymbol{y}_{k-1} = \boldsymbol{x}_{k-1} = \boldsymbol{x}_{k-2} + \dot{\boldsymbol{x}}_{k-2} + \boldsymbol{\epsilon}_{k-1}$$

and $\boldsymbol{y}_{k-2} = \boldsymbol{x}_{k-2}$. Combining these terms, we find that $\boldsymbol{y}'_k = \ddot{\boldsymbol{x}} + \boldsymbol{\epsilon}_k - \boldsymbol{\epsilon}_{k-1} + \dot{\boldsymbol{\epsilon}}_{k-1}$ which is stationary whenever $(\boldsymbol{\epsilon}_k)_{k\in\mathbb{Z}}$ and $(\dot{\boldsymbol{\epsilon}}_k)_{k\in\mathbb{Z}}$ are uncorrelated and have constant variance.