Master in Financial Engineering (EPFL)

Financial Econometrics

Exercises session 1 : Elements of correction

Exercise 1 : Characterizing second-order stationarity
Let $(\varepsilon_t)_{t\in\mathbb{Z}}$ denoted a (weak) white noise with variance $\sigma_\varepsilon^2 > 0$.

1. $X_t = \varepsilon_t - \varepsilon_{t-1}$;
   One has (for all t) :
   - Expectation

$$\begin{aligned}\mathbb{E}[X_t] &= \mathbb{E}[\varepsilon_t - \varepsilon_{t-1}] \\ &= \mathbb{E}[\varepsilon_t] - \mathbb{E}[\varepsilon_{t-1}] = 0\end{aligned}$$

   since $\mathbb{E}[\varepsilon_t] = 0$ for all t.
   - Variance

$$\begin{aligned}\mathbb{V}[X_t] &= \mathbb{V}[\varepsilon_t - \varepsilon_{t-1}] \\ &= \mathbb{V}[\varepsilon_t] + \mathbb{V}[\varepsilon_{t-1}] - 2\mathbb{C}ov(\varepsilon_t, \varepsilon_{t-1}) \qquad \text{(using the preliminary exercise)} \\ &= 2\sigma_\varepsilon^2\end{aligned}$$

   since $\mathbb{V}(\varepsilon_t) = \sigma_\varepsilon^2$ for all t, and $\mathbb{C}ov(\varepsilon_t, \varepsilon_{t-1}) = \mathbb{E}[\varepsilon_t\varepsilon_{t-1}] = 0$ (absence of correlation of a weak white noise).
   - Autocovariances
     - For $h = \pm 1$:

$$\begin{aligned}\gamma_X(1) &= \mathbb{C}ov(X_t, X_{t-1}) = \mathbb{C}ov(\varepsilon_t - \varepsilon_{t-1}, \varepsilon_{t-1} - \varepsilon_{t-2}) \\ &= -\mathbb{C}ov(\varepsilon_{t-1}, \varepsilon_{t-1}) = \mathbb{V}(\varepsilon_{t-1}) \\ &\qquad \text{(using the preliminary exercise and the absence of correlation of a WWN)} \\ &= -\sigma_\varepsilon^2.\end{aligned}$$

     - For $|h| > 1$:

$$\begin{aligned}\gamma_X(h) &= \mathbb{C}ov(X_t, X_{t-h}) = \mathbb{C}ov(\varepsilon_t - \varepsilon_{t-1}, \varepsilon_{t-h} - \varepsilon_{t-h-1}) \\ &= 0.\end{aligned}$$

   The three conditions hold—the stochastic process is covariance-stationary.

2. $X_t = a + b\varepsilon_t + c\varepsilon_{t-1}$
   One has (for all t) :
   - Expectation

$$\begin{aligned}\mathbb{E}[X_t] &= \mathbb{E}[a + b\varepsilon_t + c\varepsilon_{t-1}] \\ &= a + b\mathbb{E}[\varepsilon_t] + c\mathbb{E}[\varepsilon_{t-1}] = a\end{aligned}$$

   since $\mathbb{E}[\varepsilon_t] = 0$ for all t.
   - Variance

$$\begin{aligned}\mathbb{V}[X_t] &= \mathbb{V}[a + b\varepsilon_t + c\varepsilon_{t-1}] \\ &= b^2\mathbb{V}[\varepsilon_t] + c^2\mathbb{V}[\varepsilon_{t-1}] + 2bc\mathbb{C}ov(\varepsilon_t, \varepsilon_{t-1}) \qquad \text{(using the preliminary exercise)} \\ &= (b^2 + c^2)\sigma_\varepsilon^2\end{aligned}$$

   since $\mathbb{V}(\varepsilon_t) = \sigma_\varepsilon^2$ for all t, and $\mathbb{C}ov(\varepsilon_t, \varepsilon_{t-1}) = \mathbb{E}[\varepsilon_t\varepsilon_{t-1}] = 0$ (absence of correlation of a weak white noise).

– Autocovariances
  – For $h = \pm 1$:

$$\begin{aligned}\gamma_X(1) &= \mathbb{C}ov(X_t, X_{t-1}) = \mathbb{C}ov(a + b\varepsilon_t + c\varepsilon_{t-1}, a + b\varepsilon_{t-1} + c\varepsilon_{t-2}) \\ &= cb\mathbb{C}ov(\varepsilon_{t-1}, \varepsilon_{t-1}) \\ &\qquad \text{(using the preliminary exercise and the absence of correlation of a WWN)} \\ &= cb\sigma_\varepsilon^2.\end{aligned}$$

  – For $|h| > 1$:

$$\begin{aligned}\gamma_X(h) &= \mathbb{C}ov(X_t, X_{t-h}) = \mathbb{C}ov(a + b\varepsilon_t + c\varepsilon_{t-1}, a + b\varepsilon_{t-h} + c\varepsilon_{t-h-1}) \\ &= 0.\end{aligned}$$

The three conditions hold—the stochastic process is covariance-stationary.

3. For $t \geq 0$, $X_t - X_{t-1} = \varepsilon_t$ (one further assumes that $\forall t > 0, \varepsilon_t \perp\!\!\!\perp X_0$).
   One has :

$$\begin{aligned}X_t &= X_{t-1} + \varepsilon_t \\ &= X_{t-2} + \varepsilon_{t-1} + \varepsilon_t \\ &\vdots \\ &= X_0 + \sum_{\tau=0}^{t-1} \varepsilon_{t-\tau}.\end{aligned}$$

– Expectation

$$\begin{aligned}\mathbb{E}[X_t] &= \mathbb{E}\left[X_0 + \sum_{\tau=0}^{t-1} \varepsilon_{t-\tau}\right] \\ &= \mathbb{E}[X_0] + \sum_{\tau=0}^{t-1} \mathbb{E}[\varepsilon_{t-\tau}] \\ &= \mathbb{E}[X_0]\end{aligned}$$

for all $t > 0$. The first condition holds !

– Variance

$$\begin{aligned}\mathbb{V}[X_t] &= \mathbb{V}\left[X_0 + \sum_{\tau=0}^{t-1} \varepsilon_{t-\tau}\right] \\ &= \mathbb{V}[X_0] + \mathbb{V}\left[\sum_{\tau=0}^{t-1} \varepsilon_{t-\tau}\right] + 2\mathbb{C}ov\left[X_0, \sum_{\tau=0}^{t-1} \varepsilon_{t-\tau}\right] \\ &= \mathbb{V}[X_0] + \sum_{\tau=0}^{t-1} \mathbb{V}[\varepsilon_{t-\tau}] + 2\sum_{\tau=0}^{t-1} \mathbb{C}ov[X_0, \varepsilon_{t-\tau}] \\ &\quad \text{since} \quad \mathbb{V}\left[\sum_{\tau=0}^{t-1} \varepsilon_{t-\tau}\right] = \sum_{\tau=0}^{t-1} \mathbb{V}[\varepsilon_{t-\tau}] + \sum_{\tau} \sum_{\tau' \neq \tau} \mathbb{C}ov[\varepsilon_{t-\tau}, \varepsilon_{t-\tau'}] = \sum_{\tau=0}^{t-1} \mathbb{V}[\varepsilon_{t-\tau}] \\ &= \mathbb{V}[X_0] + t\sigma_\varepsilon^2\end{aligned}$$

since $\mathbb{V}(\varepsilon_t) = \sigma_\varepsilon^2$ for all t, and $\mathbb{C}ov(X_0, \varepsilon_{t-\tau}) = \mathbb{E}[X_0 \varepsilon_{t-\tau}] = 0$ for $\tau \in |[0, t-1]|$ ($\forall t > 0, \varepsilon_t \perp\!\!\!\perp X_0$). Therefore, the second condition does not hold. The stochastic process is nonstationary.

Exercise 2 : Linear transformation of a stationary process

1. Let (X_t) denote a weakly stationary stochastic process that has the following linear representation :

$$X_t = \mu + \sum_{k=0}^{\infty} \theta_k \epsilon_{t-k}$$

– One has :

$$\begin{aligned}
\mathbb{E}[X_t] &= \mathbb{E}\left[\mu + \sum_{k=0}^{\infty} \theta_k \epsilon_{t-k}\right] \\
&= \mu + \sum_{k=0}^{\infty} \theta_k \mathbb{E}[\epsilon_{t-k}] = \mu.
\end{aligned}$$

Moreover,

$$\begin{aligned}
\mathbb{V}[X_t] &= \mathbb{V}\left[\mu + \sum_{k=0}^{\infty} \theta_k \epsilon_{t-k}\right] = \mathbb{V}\left[\sum_{k=0}^{\infty} \theta_k \epsilon_{t-k}\right] \\
&= \sum_{k=0}^{\infty} \theta_k^2 \mathbb{V}[\epsilon_{t-k}] + \sum_{k=0}^{\infty} \sum_{j \neq k} \theta_k \theta_j \mathbb{C}ov[\epsilon_{t-k}, \epsilon_{t-j}] \\
&= \sum_{k=0}^{\infty} \theta_k^2 \mathbb{V}[\epsilon_{t-k}] \\
&\quad \text{since} \qquad \mathbb{C}ov[\epsilon_{t-k}, \epsilon_{t-j}] = \mathbb{E}[\epsilon_{t-k}\epsilon_{t-j}] = 0 \qquad \text{for } k \neq j \\
&= \sigma_\epsilon^2 \sum_{k=0}^{\infty} \theta_k^2.
\end{aligned}$$

Finally,

$$\begin{aligned}
\mathbb{C}ov[X_t, X_{t-h}] &= \mathbb{C}ov\left[\mu + \sum_{k=0}^{\infty} \theta_k \epsilon_{t-k}, \mu + \sum_{j=0}^{\infty} \theta_j \epsilon_{t-h-j}\right] \\
&= \mathbb{C}ov\left[\underbrace{\sum_{k=0}^{\infty} \theta_k \epsilon_{t-k}}_{\epsilon_t, \epsilon_{t-1}, \cdots, \epsilon_{t-h+1}, \epsilon_{t-h}, \cdots}, \underbrace{\sum_{j=0}^{\infty} \theta_j \epsilon_{t-h-j}}_{\epsilon_{t-h}, \epsilon_{t-h-1}, \cdots}\right] \\
&= \mathbb{C}ov\left[\sum_{k=0}^{h-1} \theta_k \epsilon_{t-k} + \sum_{k=h}^{\infty} \theta_k \epsilon_{t-k}, \sum_{j=0}^{\infty} \theta_j \epsilon_{t-h-j}\right] \\
&= \mathbb{C}ov\left[\sum_{k=h}^{\infty} \theta_k \epsilon_{t-k}, \sum_{j=0}^{\infty} \theta_j \epsilon_{t-h-j}\right] \\
&= \mathbb{C}ov\left[\sum_{j=0}^{\infty} \theta_{j+h} \epsilon_{t-h-j}, \sum_{j=0}^{\infty} \theta_j \epsilon_{t-h-j}\right] \qquad (\text{using } k - h = j) \\
&= \mathbb{V}\left[\sum_{j=0}^{\infty} \theta_j \epsilon_{t-h-j}\right] = \sigma_\epsilon^2 \sum_{j=0}^{\infty} \theta_j \theta_{j+h}.
\end{aligned}$$

Exercise 3 : Autocorrelation function of an autoregressive process
Consider an AR(1) process

$$X_t = \frac{4}{5} X_{t-1} + \eta_t \tag{1}$$

where η_t is a weak white noise.
Part I : ACF using the infinite moving average representation

1. By backward iteration, it is straightforward to show that [...]

$$X_t = \sum_{k=0}^{\infty} \left(\frac{4}{5}\right)^k \eta_{t-k}$$

2. Using the infinite moving average representation, one obtains :

$$\begin{aligned} \mathbb{V}[X_t] &= \mathbb{V}\left[\sum_{k=0}^{\infty} \left(\frac{4}{5}\right)^k \eta_{t-k}\right] \\ &= \sum_{k=0}^{\infty} \left(\frac{4}{5}\right)^{2k} \mathbb{V}[\eta_{t-k}] \\ &= \sigma_\epsilon^2 \sum_{k=0}^{\infty} \left(\frac{4}{5}\right)^{2k} \\ &= \frac{1}{1-0.8^2} \sigma_\eta^2 \end{aligned}$$

since (η_t) is a weak white noise and is even the innovation process of (X_t).

3. The derivation of the autocorrelation function is straightforward (using Exercise 2 or the Yule-Walker equation). One obtains :

$$\rho_X(k) = 0.8^k.$$

Part II : ACF using the Yule-Walker equation

1. Multiplying Eq. (1) by X_t and taking the expectation on both sides yield :

$$\gamma_X(0) = 0.8\gamma_X(1) + \sigma_\eta^2.$$

On the other hand, multiplying Eq. (1) by X_{t-1} and taking the expectation on both sides yield :

$$\gamma_X(1) = 0.8\gamma_X(0).$$

Using the last two equations, one obtains $\gamma_X(0)$ (variance) and the autocovariance of order 1, and thus the autocorrelation of order 1. More generally, for $|h| \geq 1$, this is the so-called Yule-Walker equation for the autocovariance (respectively, autocorrelation function).

2. Show that the autocorrelation function is driven by a difference equation of order 1 : See previous question.
3. Solving this difference equation provides the same results as those of Part I!

Part III : PACF

1. By definition!
2. The affirmation of the student is correct. Indeed, if the higher order partial autocorrelation are different from zero, it means that the process is not an AR(1) (and is thus misspecified).

Part IV : Extensions

1. Consider the following AR(2) process

$$X_t = \frac{1}{3} + \frac{5}{6}X_{t-1} - \frac{1}{6}X_{t-2} + \eta_t$$

where η_t is a weak white noise.
One first demeans the stochastic process. In so doing, it is worth noting that (since (X_t) is weakly stationary)

$$\mathbb{E}[X_t] = \frac{1/3}{1 - 5/6 + 1/6} = 1.$$

Let $\tilde{X}_t$ denote (for all t)

$$\tilde{X}_t = X_t - 1.$$

It follows that $(\tilde{X}_t, t \in \mathbb{Z})$ is defined by :

$$\tilde{X}_t = \frac{5}{6}\tilde{X}_{t-1} - \frac{1}{6}\tilde{X}_{t-2} + \eta_t.$$

Notably, it is straightforward to show that the autocorrelation function of (X_t) is exactly the same as the one of $(\tilde{X}_t)$. Then,

- Step 1 : Multiply both sides by $\tilde{X}_t$ and take the expectation :

$$\gamma_{\tilde{X}}(0) = \frac{5}{6}\gamma_{\tilde{X}}(1) - \frac{1}{6}\gamma_{\tilde{X}}(2) + \sigma_\eta^2.$$

- Step 2 : Multiply both sides by $\tilde{X}_{t-1}$ and take the expectation :

$$\gamma_{\tilde{X}}(1) = \frac{5}{6}\gamma_{\tilde{X}}(0) - \frac{1}{6}\gamma_{\tilde{X}}(1).$$

- Step 3 : Multiply both sides by $\tilde{X}_{t-2}$ and take the expectation :

$$\gamma_{\tilde{X}}(2) = \frac{5}{6}\gamma_{\tilde{X}}(1) - \frac{1}{6}\gamma_{\tilde{X}}(0).$$

  Using the last three equations, one gets the variance, the autocovariance of order 1 (and the autocovariance of order 2).

- Step 4 : More generally, for $|h| \geq 2$:

$$\gamma_{\tilde{X}}(h) = \frac{5}{6}\gamma_{\tilde{X}}(h-1) - \frac{1}{6}\gamma_{\tilde{X}}(h-2)$$

  and

$$\rho_{\tilde{X}}(h) = \frac{5}{6}\rho_{\tilde{X}}(h-1) - \frac{1}{6}\rho_{\tilde{X}}(h-2).$$

  The autocovariance (respectively, autocorrelation) function is defined from a difference equation of order 2 (i.e., one needs two initial conditions if the two roots are distinct).

Exercise 4 : The ARMA(1,1) is defined to be

$$X_t = \frac{1}{3} + \frac{1}{8}X_{t-1} + \epsilon_t - \frac{3}{4}\epsilon_{t-1}$$

where (ϵ_t) is a weak white noise.
Before answering questions 1-3, we first demean the stochastic process. In so doing, it is worth noting that (since (X_t) is weakly stationary)

$$\mathbb{E}\left[X_t\right] = \frac{1/3}{1-1/8} = \frac{8}{21}.$$

Let $\tilde{X}_t$ denote (for all t)

$$\tilde{X}_t = X_t - \frac{8}{21}.$$

It follows that $(\tilde{X}_t, t \in \mathbb{Z})$ is defined by :

$$\tilde{X}_t = \frac{1}{8}\tilde{X}_{t-1} + \epsilon_t - \frac{3}{4}\epsilon_{t-1}. \tag{2}$$

Notably, it is straightforward to show that the autocorrelation function of (X_t) is exactly the same as the one of $(\tilde{X}_t)$.

1. The moving average process is defined by (for all t) :

$$W_t = \epsilon_t - \frac{3}{4}\epsilon_{t-1}.$$

It is straightforward to show that

$$\rho_W(h) = \begin{cases} 1 & \text{if } h = 0 \\ \frac{-3/4}{1+9/16} & \text{if } h = \pm 1 \\ 0 & \text{if } |h| > 1. \end{cases}$$

2. For $h = 0$, we can multiply Eq. (2) by $\tilde{X}_t$ and take the expectation on both sides :

$$\mathbb{E}\left[\tilde{X}_t^2\right] = \frac{1}{8}\mathbb{E}\left[\tilde{X}_t\tilde{X}_{t-1}\right] + \mathbb{E}\left[\tilde{X}_t\epsilon_t\right] - \frac{3}{4}\mathbb{E}\left[\tilde{X}_t\epsilon_{t-1}\right].$$

Therefore, one has

$$\gamma_{\tilde{X}}(0) = \frac{1}{8}\gamma_{\tilde{X}}(1) + \sigma_\epsilon^2 - \frac{3}{4}\mathbb{E}\left[\tilde{X}_t\epsilon_{t-1}\right]$$

where

$$\begin{aligned} \mathbb{E}\left[\tilde{X}_t\epsilon_{t-1}\right] &= \mathbb{E}\left[\left(\frac{1}{8}\tilde{X}_{t-1} + \epsilon_t - \frac{3}{4}\epsilon_{t-1}\right)\epsilon_{t-1}\right] \\ &= \frac{1}{8}\mathbb{E}\left[\tilde{X}_{t-1}\epsilon_{t-1}\right] - \frac{3}{4}\mathbb{E}\left[\epsilon_{t-1}^2\right] \\ &= \left(\frac{1}{8} - \frac{3}{4}\right)\sigma_\epsilon^2 = -\frac{5}{8}\sigma_\epsilon^2. \end{aligned}$$

In the same respect (after multiplying by $\tilde{X}_{t-1}$ and taking the expectation on both sides), for $h = 1$, one gets

$$\gamma_{\tilde{X}}(1) = \frac{1}{8}\gamma_{\tilde{X}}(0) - \frac{3}{4}\mathbb{E}\left[\tilde{X}_{t-1}\epsilon_{t-1}\right]$$

which is equivalent to

$$\gamma_{\tilde{X}}(1) = \frac{1}{8}\gamma_{\tilde{X}}(0) - \frac{3}{4}\sigma_\epsilon^2.$$

Using the two equations ($h = 0$ and 1), one can obtain the autocovariance of order 0 (i.e., the variance) and the autocovariance of order 1 (and thus the autocorrelation of order 1).

3. When $|h| \geq 2$, the moving average part does not contribute (due to the fact that (ϵ_t) is the innovation process). Consequently, the Yule-Walker equation (for the autocovariance function) is given by

$$\gamma_{\tilde{X}}(h) = \frac{1}{8}\gamma_{\tilde{X}}(h-1)$$

and the autocorrelation function is given by a difference equation of order one :

$$\rho_{\tilde{X}}(h) = \frac{1}{8}\rho_{\tilde{X}}(h-1).$$

Exercise 5 Consider the following AR(p) stochastic processes

(i) $X_t = \frac{1}{2} + \frac{4}{5}X_{t-1} + \epsilon_t$, where ϵ_t is a weak white noise $(0, \sigma_\epsilon^2)$.

(a) First, write this stochastic process in mean-deviation

$$\tilde{X}_t = \frac{4}{5}\tilde{X}_{t-1} + \epsilon_t$$

where $\tilde{X}_t = X_t - m_X$ with $m_X = \frac{\frac{1}{2}}{1-\frac{4}{5}} = \frac{5}{2}$.

The best linear forecast for $h = 1$ is defined to be

$$\tilde{X}_t^*(1) = \mathbb{EL}\left[\tilde{X}_{t+1} \mid \tilde{X}_t, \tilde{X}_{t-1}, \cdots\right]$$

where

$$\tilde{X}_{t+1} = \frac{4}{5}\tilde{X}_t + \epsilon_{t+1}.$$

Since the representation of (X_t) is fundamental, (ϵ_t) is the innovation process of (X_t) (and $(\tilde{X}_t)$) and

$$\mathbb{EL}\left[\epsilon_{t+1} \mid \tilde{X}_t, \tilde{X}_{t-1}, \cdots\right] = 0.$$

Therefore

$$\begin{aligned} \tilde{X}_t^*(1) &= \mathbb{EL}\left[\frac{4}{5}\tilde{X}_t + \epsilon_{t+1} \mid \tilde{X}_t, \tilde{X}_{t-1}, \cdots\right] \\ &= \frac{4}{5}\tilde{X}_t \end{aligned}$$

and

$$\begin{aligned} X_t^*(1) &= \tilde{X}_t^*(1) + m_X \\ &= m_X + \frac{4}{5}(X_t - m_X) \\ &= \frac{1}{2} + \frac{4}{5}X_t. \end{aligned}$$

The best linear forecast for $h = 2$ is defined to be

$$\tilde{X}_t^*(2) = \mathbb{EL}\left[\tilde{X}_{t+2} \mid \tilde{X}_t, \tilde{X}_{t-1}, \cdots\right]$$

where

$$\tilde{X}_{t+2} = \frac{4}{5}\tilde{X}_{t+1} + \epsilon_{t+2}.$$

Therefore

$$\begin{aligned}
\tilde{X}_t^*(2) &= \mathbb{EL}\left[\frac{4}{5}\tilde{X}_{t+1} + \epsilon_{t+2} \mid \tilde{X}_t, \tilde{X}_{t-1}, \cdots\right] \\
&= \frac{4}{5}\mathbb{EL}\left[\tilde{X}_{t+1} \mid \tilde{X}_t, \tilde{X}_{t-1}, \cdots\right] + \mathbb{EL}\left[\epsilon_{t+2} \mid \tilde{X}_t, \tilde{X}_{t-1}, \cdots\right] \\
&= \frac{4}{5}X_t^*(1) \\
&= \left(\frac{4}{5}\right)^2 \tilde{X}_t.
\end{aligned}$$

and

$$\begin{aligned}
X_t^*(2) &= \tilde{X}_t^*(2) + m_X \\
&= \left(\frac{4}{5}\right)^2 \tilde{X}_t + m_X \\
&= \left(\frac{4}{5}\right)^2 X_t + m_X\left(1 - \left(\frac{4}{5}\right)^2\right).
\end{aligned}$$

(b) The forecast error for $h = 1$ is given by

$$\begin{aligned}
e_t(1) &= X_{t+1} - X_t^*(1) \equiv \tilde{X}_{t+1} - \tilde{X}_t^*(1) \\
&= \frac{4}{5}\tilde{X}_t + \epsilon_{t+1} - \frac{4}{5}\tilde{X}_t \\
&= \epsilon_{t+1}
\end{aligned}$$

and the corresponding variance is

$$\mathbb{V}\left[e_t(1)\right] = \mathbb{V}\left[\epsilon_{t+1}\right] = \sigma_\epsilon^2.$$

The forecast error for $h = 2$ is given by [1]

$$\begin{aligned}
e_t(2) &= X_{t+2} - X_t^*(2) \equiv \tilde{X}_{t+2} - \tilde{X}_t^*(2) \\
&= \frac{4}{5}\tilde{X}_{t+1} + \epsilon_{t+2} - \frac{4}{5}\tilde{X}_t^*(1) \\
&= \frac{4}{5}\left(\tilde{X}_{t+1} - \tilde{X}_t^*(1)\right) + \epsilon_{t+2} \\
&= \frac{4}{5}e_t(1) + \epsilon_{t+2} \\
&= \frac{4}{5}\epsilon_{t+1} + \epsilon_{t+2}
\end{aligned}$$

1. One can also write

$$\begin{aligned}
e_t(2) &= X_{t+2} - X_t^*(2) \equiv \tilde{X}_{t+2} - \tilde{X}_t^*(2) \\
&= \frac{4}{5}\tilde{X}_{t+1} + \epsilon_{t+2} - \frac{4}{5}\tilde{X}_t^*(1) \\
&= \frac{4}{5}\left(\frac{4}{5}\tilde{X}_t + \epsilon_{t+1}\right) + \epsilon_{t+2} - \left(\frac{4}{5}\right)^2 \tilde{X}_t \\
&= \frac{4}{5}\epsilon_{t+1} + \epsilon_{t+2}.
\end{aligned}$$

and the corresponding variance is

$$\mathbb{V}\left[e_t(2)\right] = \mathbb{V}\left[\frac{4}{5}\epsilon_{t+1} + \epsilon_{t+2}\right] = \left[1 + \left(\frac{4}{5}\right)^2\right]\sigma_\epsilon^2.$$

(c) More generally, the h-step ahead forecast (for $h \geq 1$) is defined to be

$$\tilde{X}_t^*(h) = \mathbb{EL}\left[\tilde{X}_{t+h} \mid \tilde{X}_t, \tilde{X}_{t-1}, \cdots\right]$$

where

$$\tilde{X}_{t+h} = \frac{4}{5}\tilde{X}_{t+h-1} + \epsilon_{t+h}.$$

Using the fundamental representation and the linear expectation properties, one gets

$$\begin{aligned}
\tilde{X}_t^*(h) &= \mathbb{EL}\left[\frac{4}{5}\tilde{X}_{t+h-1} + \epsilon_{t+h} \mid \tilde{X}_t, \tilde{X}_{t-1}, \cdots\right] \\
&= \frac{4}{5}\mathbb{EL}\left[\tilde{X}_{t+h-1} \mid \tilde{X}_t, \tilde{X}_{t-1}, \cdots\right] + \mathbb{EL}\left[\epsilon_{t+h} \mid \tilde{X}_t, \tilde{X}_{t-1}, \cdots\right] \\
&= \frac{4}{5}X_t^*(h-1).
\end{aligned}$$

This is a difference equation of order one. The characteristic equation is the same as the one used for the stability condition (the fundamentalness of the representation).

On the other hand, the forecast error is defined to be

$$\begin{aligned}
e_t(h) &= X_{t+h} - X_t^*(h) \equiv \tilde{X}_{t+h} - \tilde{X}_t^*(h) \\
&= \frac{4}{5}\tilde{X}_{t+h-1} + \epsilon_{t+h} - \frac{4}{5}\tilde{X}_t^*(h-1) \\
&= \frac{4}{5}\left(\tilde{X}_{t+h-1} - \tilde{X}_t^*(h-1)\right) + \epsilon_{t+h} \\
&= \frac{4}{5}e_t(h-1) + \epsilon_{t+h}.
\end{aligned}$$

Using a backward induction, one gets [2]

$$\begin{aligned}
e_t(h) &= \epsilon_{t+h} + \left(\frac{4}{5}\right)\epsilon_{t+h-1} + \cdots + \left(\frac{4}{5}\right)^{h-1}\epsilon_{t+1} \\
&\quad \text{(since } e_t(h) = 0 \text{ for } h \leq 0\text{)} \\
&= \sum_{k=0}^{h-1}\left(\frac{4}{5}\right)^k \epsilon_{t+h-k}.
\end{aligned}$$

2. One can also the infinite moving average representation :

$$\tilde{X}_{t+h} = \sum_{k=0}^{\infty}\left(\frac{4}{5}\right)^k \epsilon_{t+h-k}$$

and the equivalent definition of the best linear h-step ahead forecast

$$\tilde{X}_t^*(h) = \sum_{k=h}^{\infty}\left(\frac{4}{5}\right)^k \epsilon_{t+h-k}.$$

It follows that

$$\begin{aligned}
e_t(h) &= X_{t+h} - X_t^*(h) \equiv \tilde{X}_{t+h} - \tilde{X}_t^*(h) \\
&= \sum_{k=0}^{\infty}\left(\frac{4}{5}\right)^k \epsilon_{t+h-k} - \sum_{k=h}^{\infty}\left(\frac{4}{5}\right)^k \epsilon_{t+h-k} \\
&= \sum_{k=0}^{h-1}\left(\frac{4}{5}\right)^k \epsilon_{t+h-k}.
\end{aligned}$$

The variance of the h-step ahead forecast is given by :

$$\begin{aligned}\mathbb{V}\left[e_t(h)\right] &= \mathbb{V}\left[\sum_{k=0}^{h-1}\left(\frac{4}{5}\right)^k \epsilon_{t+h-k}\right] \\ &= \sigma_\epsilon^2 \sum_{k=0}^{h-1}\left(\frac{4}{5}\right)^{2k} \\ &= \sigma_\epsilon^2 \frac{1-\left(\frac{4}{5}\right)^{2h}}{1-\left(\frac{4}{5}\right)^2}.\end{aligned}$$

(d) It follows that

$$\begin{aligned}\tilde{X}_t^*(h) &= \left(\frac{4}{5}\right)^h \tilde{X}_t \\ &\underset{h\to\infty}{\longrightarrow} 0 \equiv \mathbb{E}\left[\tilde{X}_t\right]\end{aligned}$$

and

$$\begin{aligned}X_t^*(h) &= \tilde{X}_t^*(h) + m_X \\ &\underset{h\to\infty}{\longrightarrow} m_X.\end{aligned}$$

(ii) $X_t = \frac{5}{6}X_{t-1} - \frac{1}{6}X_{t-2} + \epsilon_t$, where ϵ_t is a weak white noise $(0, \sigma_\epsilon^2)$.

(a) The best linear forecast for $h = 1$ is defined to be

$$X_t^*(1) = \mathbb{EL}\left[X_{t+1} \mid X_t, X_{t-1}, \cdots\right]$$

where

$$X_{t+1} = \frac{5}{6}X_t - \frac{1}{6}X_{t-1} + \epsilon_{t+1}.$$

Using the fundamentalness of the representation (the two roots of the characteristic equation, $\frac{1}{2}$ and $\frac{1}{3}$, are of modulus less than one) and the properties of $\mathbb{EL}$, it follows that

$$\begin{aligned}X_t^*(1) &= \mathbb{EL}\left[\frac{5}{6}X_t - \frac{1}{6}X_{t-1} + \epsilon_{t+1} \mid X_t, X_{t-1}, \cdots\right] \\ &= \mathbb{EL}\left[\frac{5}{6}X_t - \frac{1}{6}X_{t-1} \mid X_t, X_{t-1}, \cdots\right] + \mathbb{EL}\left[\epsilon_{t+1} \mid X_t, X_{t-1}, \cdots\right] \\ &= \frac{5}{6}X_t - \frac{1}{6}X_{t-1}.\end{aligned}$$

The best linear forecast for $h = 2$ is defined to be

$$X_t^*(2) = \mathbb{EL}\left[X_{t+2} \mid X_t, X_{t-1}, \cdots\right]$$

where

$$X_{t+2} = \frac{5}{6}X_{t+1} - \frac{1}{6}X_t + \epsilon_{t+2}.$$

Therefore

$$\begin{aligned}X_t^*(2) &= \mathbb{EL}\left[\frac{5}{6}X_{t+1} - \frac{1}{6}X_t + \epsilon_{t+2} \mid X_t, X_{t-1}, \cdots\right] \\ &= \mathbb{EL}\left[\frac{5}{6}X_{t+1} - \frac{1}{6}X_t \mid X_t, X_{t-1}, \cdots\right] + \mathbb{EL}\left[\epsilon_{t+2} \mid X_t, X_{t-1}, \cdots\right] \\ &= \frac{5}{6}X_t^*(1) - \frac{1}{6}X_t.\end{aligned}$$

(b) The forecast error of the one-step ahead forecast is given by

$$e_t(1) = X_{t+1} - X_t^*(1) = \epsilon_{t+1}$$

and

$$\mathbb{V}\left[e_t(1)\right] = \mathbb{V}\left[\epsilon_{t+1}\right] = \sigma_\epsilon^2.$$

The forecast error of the two-step ahead forecast is given by

$$\begin{aligned} e_t(2) &= X_{t+2} - X_t^*(2) \\ &= \frac{5}{6}X_{t+1} - \frac{1}{6}X_t + \epsilon_{t+2} - \frac{5}{6}X_t^*(1) + \frac{1}{6}X_t \\ &= \frac{5}{6}\left(X_{t+1} - X_t^*(1)\right) + \epsilon_{t+2} \\ &= \frac{5}{6}e_t(1) + \epsilon_{t+2} \\ &= \frac{5}{6}\epsilon_{t+1} + \epsilon_{t+2}. \end{aligned}$$

and

$$\mathbb{V}\left[e_t(2)\right] = \mathbb{V}\left[\frac{5}{6}\epsilon_{t+1} + \epsilon_{t+2}\right] = \left[1 + \left(\frac{5}{6}\right)^2\right]\sigma_\epsilon^2.$$

(c) The best h-step ahead forecast is given by

$$X_t^*(h) = \mathbb{EL}\left[X_{t+h} \mid X_t, X_{t-1}, \cdots\right]$$

where

$$X_{t+h} = \frac{5}{6}X_{t+h-1} - \frac{1}{6}X_{t+h-2} + \epsilon_{t+h}.$$

Therefore, for $h \geq 2$

$$\begin{aligned} X_t^*(h) &= \mathbb{EL}\left[\frac{5}{6}X_{t+h-1} - \frac{1}{6}X_{t+h-2} + \epsilon_{t+h} \mid X_t, X_{t-1}, \cdots\right] \\ &= \frac{5}{6}\mathbb{EL}\left[X_{t+h-1} \mid X_t, X_{t-1}, \cdots\right] - \frac{1}{6}\mathbb{EL}\left[X_{t+h-2} \mid X_t, X_{t-1}, \cdots\right] + \mathbb{EL}\left[\epsilon_{t+h} \mid X_t, X_{t-1}, \cdots\right] \\ &= \frac{5}{6}X_t^*(h-1) - \frac{1}{6}X_t^*(h-2). \end{aligned}$$

This is a difference equation of order 2—the roots of the characteristic equation are the ones obtained for the fundamental representation. The general solution is

$$X_t^*(h) = A\left(\frac{1}{3}\right)^h + B\left(\frac{1}{2}\right)^h$$

where A and B can be determined by using two initial conditions.

(d) As $h \to \infty$, one has

$$X_t^*(h) \underset{h\to\infty}{\longrightarrow} 0.$$

Exercise 6 Let (X_t) denote the following stochastic process

$$X_t = \frac{1}{3} + \epsilon_t - \frac{3}{4}\epsilon_{t-1} + \frac{1}{8}\epsilon_{t-2}$$

1. The best h-step ahead forecast is given by

$$X_t^*(h) = \mathbb{EL}\left[X_{t+h} \mid X_t, X_{t-1}, \cdots\right]$$

where

$$X_{t+h} = \frac{1}{3} + \epsilon_{t+h} - \frac{3}{4}\epsilon_{t+h-1} + \frac{1}{8}\epsilon_{t+h-2}.$$

Therefore (for $h > 0$)

$$\begin{aligned}
X_t^*(h) &= \mathbb{EL}\left[\frac{1}{3} + \epsilon_{t+h} - \frac{3}{4}\epsilon_{t+h-1} + \frac{1}{8}\epsilon_{t+h-2} \mid X_t, X_{t-1}, \cdots\right] \\
&= \frac{1}{3} + \mathbb{EL}\left[\epsilon_{t+h} \mid X_t, X_{t-1}, \cdots\right] - \frac{3}{4}\mathbb{EL}\left[\epsilon_{t+h-1} \mid X_t, X_{t-1}, \cdots\right] + \frac{1}{8}\mathbb{EL}\left[\epsilon_{t+h-2} \mid X_t, X_{t-1}, \cdots\right] \\
&= \frac{1}{3} + \mathbb{EL}\left[\epsilon_{t+h} \mid \epsilon_t, \epsilon_{t-1}, \cdots\right] - \frac{3}{4}\mathbb{EL}\left[\epsilon_{t+h-1} \mid \epsilon_t, \epsilon_{t-1}, \cdots\right] + \frac{1}{8}\mathbb{EL}\left[\epsilon_{t+h-2} \mid \epsilon_t, \epsilon_{t-1}, \cdots\right] \\
&= \frac{1}{3} - \frac{3}{4}\tilde{\epsilon}_{t+h-1} + \frac{1}{8}\tilde{\epsilon}_{t+h-2}
\end{aligned}$$

where

$$\tilde{\epsilon}_{t+h-k} = \begin{cases} 0 & \text{if } h > k \\ \epsilon_{t+h-k} & \text{if } h \leq k. \end{cases}$$

Finally,

- $h = 1$

$$X_t^*(1) = \frac{1}{3} - \frac{3}{4}\epsilon_t + \frac{1}{8}\epsilon_{t-1}$$

- $h = 2$

$$X_t^*(2) = \frac{1}{3} + \frac{1}{8}\epsilon_t$$

- $h > 2$

$$X_t^*(h) = \frac{1}{3}.$$

Remark : The first two forecasts $X_t^*(1)$ and $X_t^*(2)$ cannot be used in practise (since ϵ_t and ϵ_{t-1} are not observable). How can one proceed ?...

2. The forecast error is given by

$$\begin{aligned}
e_t(h) &= X_{t+h} - X_t^*(h) \\
&= \frac{1}{3} + \epsilon_{t+h} - \frac{3}{4}\epsilon_{t+h-1} + \frac{1}{8}\epsilon_{t+h-2} \\
&\quad -\frac{1}{3} + \frac{3}{4}\tilde{\epsilon}_{t+h-1} - \frac{1}{8}\tilde{\epsilon}_{t+h-2} \\
&= \epsilon_{t+h} - \frac{3}{4}\left(\epsilon_{t+h-1} - \tilde{\epsilon}_{t+h-1}\right) + \frac{1}{8}\left(\epsilon_{t+h-2} - \tilde{\epsilon}_{t+h-2}\right)
\end{aligned}$$

where

$$\tilde{\epsilon}_{t+h-k} = \begin{cases} 0 & \text{if } h > k \\ \epsilon_{t+h-k} & \text{if } h \leq k. \end{cases}$$

Therefore

- $h = 1$

$$e_t(1) = \epsilon_{t+1}$$

and

$$\mathbb{V}\left[e_t(1)\right] = \mathbb{V}\left[\epsilon_{t+1}\right] = \sigma_\epsilon^2.$$

- $h = 2$

$$e_t(2) = \epsilon_{t+2} - \frac{3}{4}\epsilon_{t+1}$$

and

$$\begin{aligned}\mathbb{V}\left[e_t(2)\right] &= \mathbb{V}\left[\epsilon_{t+2} - \frac{3}{4}\tilde{\epsilon}_{t+1}\right] \\ &= \left[1 + \left(\frac{3}{4}\right)^2\right]\sigma_\epsilon^2.\end{aligned}$$

- $h > 2$

$$e_t(h) = \epsilon_{t+h} - \frac{3}{4}\epsilon_{t+h-1} + \frac{1}{8}\epsilon_{t+h-2}$$

and

$$\begin{aligned}\mathbb{V}\left[e_t(h)\right] &= \mathbb{V}\left[\epsilon_{t+h} - \frac{3}{4}\epsilon_{t+h-1} + \frac{1}{8}\epsilon_{t+h-2}\right] \\ &= \left[1 + \left(\frac{3}{4}\right)^2 + \left(\frac{1}{8}\right)^2\right]\sigma_\epsilon^2.\end{aligned}$$

Exercise 7 : Consider the interest rate spread variable over the period 1960Q1-2010Q1.

1. Some elements :
   - Using Figure 1, it is difficult to assess (visual procedure) whether or not the series is stationary. All in all, the series might be mean-reverting but with a large degree of persistence.
   - Using Figure 2 (acf of X_t), the decreasing rate of the autocorrelation function (only five lags are statistically different from zero using a Bartlett's correction for the confidence bands) tends to favor the assumption that the series may be (weakly) stationary. Figure 2 also suggest that $q_{\max} = 5$.
   - Using Figure 3, one can identify a upper bound for the autoregressive part (as long as the series is weakly stationary). One can choose $p_{\max} = 9$. Using the parsimony principle (in a first step) will rather suggest $p_{\max} = 6$.
   - Using Figure 4 and Figure 5, one can identify a upper bound for the moving average part (respectively, autoregressive part) if the series X_t is (asymptotically) stationary in first-difference...
2. Different unit root tests are conducted in order to identify d.
   (a) The `DF` unit root test is presented in Table 1.

Table 1 : `DF` unit root test

Parameter	Estimation	Std. Err.	t-stat.	p-value
ϕ	-0.1111	0.0330	-3.3624	0.0009
c	0.1575	0.0603	2.6083	0.0098

Note : Tabulated critical values of the Student test statistic at 1%, 5% et 10% are respectively -3.464, -2.876 and -2.574,.

   - There is no apparent trend in Figure 1. Therefore, case 2 makes more sense than case 4 (see handouts). The test regression is

$$\Delta X_t = \phi X_{t-1} + c + \epsilon_t$$

   The null hypothesis is $H_0 : \phi = 0$. The alternative hypothesis is $H_a : \phi < 0$.
   - The `DF` Student test statistic is -3.3624. [3]
   - The test statistic falls below the 1% critical value (or the test statistic is not greater than the critical values) : one rejects the null hypothesis of nonstationarity. [4]

   (b) Table 2 provides the results of the `ADF` unit root test using `SBIC` (the maximum number of lags is 12).

Table 2 : `ADF` unit root test

Parameter	Estimation	Std. Err.	t-stat.	p-value
	A. One lag			
$\phi^*(1)$	-0.1362	0.0331	-4.1183	0.0001
α_1	0.2454	0.7079	3.4663	0.0007
c	0.1890	0.0598	3.1588	0.0018

Note : Tabulated critical values of the Student test statistic at 1%, 5% et 10% are respectively -3.464, -2.876 and -2.574,.

   - The optimal number of lags is one (using the `SBIC`). The test regression is given by

$$\Delta X_t = \phi^*(1) X_{t-1} + \alpha_1 \Delta X_{t-1} + \epsilon_t$$

   The null hypothesis is $H_0 : \phi^*(1) = 0$. The alternative hypothesis is $H_a : \phi^*(1) < 0$. The t_{ADF} statistic, which is -4.1183, falls below the 1% (respectively, 5%, 10%) critical value : one rejects the null hypothesis of nonstationarity.

3. It is worth noting that one cannot use the p-value to interpret this test-statistic since the asymptotic distribution is nonstandard !

4. The test is a left one-sided.

(c) Table 3 provides the results of the PP unit root test.

Table 1 : DF unit root test

Parameter	Estimation	Std. Err.	t-stat.	p-value
ϕ	-0.1111	0.0330	-3.3624	0.0009
c	0.1575	0.0603	2.6083	0.0098

Note : Tabulated critical values of the Student test statistic at 1%, 5% et 10% are respectively -3.464, -2.876 and -2.574,.

The PP test statistic is given by -3.7910.

- The PP test implements a nonparametric correction of the DF test statistics. In so doing, the test regression is the same and thus the estimates do not change (see Lecture notes).
- Again...one rejects the null of nonstationarity (the interpretation is the same as in the Dickey-Fuller test).

(d) Finally, a KPSS unit root test (with only a constant) is conducted. The LM test statistic is given by 0.2959. The asymptotic critical values are given by 0.739 (1% level), 0.463 (5% level), and 0.347 (10% level).

The KPSS statistic is lower than the critical value at 1%, 5% or 10% : one cannot reject the null of stationarity. In this respect, the two types of test (null of stationarity and null of nonstationarity) lead to the same conclusion.

(e) One might choose $d = 0$.

3. Using the autocorrelation function (Figure 2) and the partial autocorrelation function (Figure 3), determine some orders p and q. *See Question 1.*
4. Table 4 reports the information criteria AIC (panel A), SBIC (panel B), and HQ (panel C). Some comments
   - Choose the models that minimize each information criterion (the lowest value as well as adjacent values).
   - For example, using SBIC, one may choose ARMA()(1,1), ARMA()(2,1), etc.
   - Here, choosing $q > 5$ is not a good strategy, especially for AIC, given the order identification ($q_{\max} = 5$).[5]
5. Some comments
   - The two specifications ARMA()(2,6) and ARMA()(2,7) are quite close in terms of information criteria (AIC, SBIC or HQ).[6] At the same time, the estimates are completely different...While all estimates are statistically different from zero (with the exception of ϕ_1) in the case of an ARMA()(2,6), all estimates are not statistically different from zero (with the exception of the constant term) in the case of an ARMA()(2,7).[7]
   - In contrast, the ARMA()(2,(1,7)) specification, which is defined by

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_7 \epsilon_{t-7},$$

   performs very well in terms of in-sample information criteria. This specification is obtained by imposing some linear constraints on the ARMA()(2,7). The choice of this specification is rationalized in Question 6 (autocorrelation tests). At the end, this specification has the lowest value for the SBIC (this is not true for the AIC...but the AIC may over-parameterize in this application...).
6. Table 7 displays the (sample) autocorrelation function and the (sample) partial autocorrelation function. Moreover, the Portmanteau test is implemented for each model of the previous question.

5. It is worth noting that the values of the information criteria provide some relevant information regarding the specification as long as this information is correctly used. See further.

6. Compare the values in Table 4 !

7. The fact that the standard errors increase substantially after adding a lag suggests that the information regarding the moving average part is redundant and that there is a problem of pseudo-colinearity.

– It is useful to look at the (sample) autocorrelation function of the residuals in order to check that the residuals behave like a (weak) white noise (as assumed in the definition). In the presence of autocorrelation, the retained model (step 3 : Box and Jenkins) is misspecified (and one may go back to step 1 or step 2 to improve the specification).
– The Box-Pierce Q-statistic (or portmanteau test) tests the joint hypothesis that the first K autocorrelations of the adjusted error terms are jointly zero :

$$H_0 : \rho_{\hat{\epsilon}}(1) = \rho_{\hat{\epsilon}}(2) = \cdots = \rho_{\hat{\epsilon}}(K) = 0.$$

The test statistic is given by :

$$Q = T \sum_{k=1}^{K} \hat{\rho}_{\hat{\epsilon}}^2(k)$$

where $\hat{\rho}_{\hat{\epsilon}}^2(k)$ is the k-th order sample autocorrelation of the estimated residuals, T is the sample size, and K is chosen sufficiently large.
The Q-test has an asymptotic chi-square (χ^2) distribution with $K-p-q$ degrees of freedom.
– The null hypothesis of uncorrelated (estimated) residuals is rejected if the observed test statistic, Q, exceeds the tabulated critical value (for a chosen significance level). The null hypothesis is rejected in the following cases : AR()(2), ARMA()(1,1), ARMA()(2,1). In the case of an ARMA()(2,1), the null hypothesis is rejected for K large enough. The null hypothesis (absence of correlation) is not rejected for the AR()(7) and ARMA()(2,(1,7)). [8] It is worth noticing that the autocorrelation of order 7 is generally large when the null hypothesis is rejected : this tends to justify why an ARMA()(2,(1,7)) is estimated (in order to capture this autocorrelation of the residuals, one can add a lag of order 7 in the moving average part).
– Choose AR()(7) and ARMA()(2,(1,7)) !

7. Finally, one-step ahead forecasts are implemented : the model is estimated with a recursive window. The first estimation is done over the period 1960Q4-1995Q3. In order to compare the one-step ahead forecasts, one compute the Diebold-Mariano test using the `RMSE` (respectively, `MAE`).
   – See lecture notes.
   – Some comments
     – Using `RMSE`, one cannot reject the null hypothesis of equal accuracy.
     – Using `MAE`, the ARMA()(1,1) specification "outperforms" other models (for $h = 1$). Interestingly, this model was misspecified : in-sample performances are not the same as out-of-sample performances !
     – All in all, the conclusion depends on the out-of-sample information criterion. It may also depend on the forecasting horizon (only $h = 1$ is considered !). [9]
8. Bonus question ! As a final check, one computes the following regression

$$X_t = a + bX_{t-1}^*(1) + u_t$$

for all t in the holdout period.

Some comments :
– Using this specification, one should expect that the estimate of b is close to one (why ?) and that the estimate of a is not statistically different from zero (otherwise, it means that there is a systematic bias in the forecasts).
– Table 10 shows that one cannot reject the null hypothesis that a equals zero, etc. Interestingly, the two models, AR()(7) and ARMA()(2,(1,7)), perform better than the others...

8. It is worth noting that the first $p+q$ statistics and p-values do not make sense since the asymptotic distribution is not defined.
9. Other issues are the finite sample behavior of this test, the correction for autocorrelation, etc.