FORECASTING

Anwsers to exercises 1-14

Guillaume Chevillon

March 8, 2017

EXERCISE 1

We only need developing

$$\begin{aligned}
\mathsf{Var}[X] &= \mathsf{E}\left[(X-\mathsf{E}[X])^2\right] \\
&= \mathsf{E}\left[X^2 - 2X\mathsf{E}[X] + \mathsf{E}[X]^2\right] \\
&= \mathsf{E}\left[X^2\right] - 2\mathsf{E}[X]\,\mathsf{E}[X] + \mathsf{E}[X]^2 \\
&= \mathsf{E}\left[X^2\right] - \mathsf{E}[X]^2
\end{aligned}$$

and similarly with the covariance. What is important is that it also works with sample means, replacing $\mathsf{E}[X]$ with

$$\overline{X}_n = n^{-1}\sum_{i=1}^{n} X_i$$

and $\mathsf{E}\left[(X-\mathsf{E}[X])^2\right]$ with

$$n^{-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2.$$

EXERCISE 2

Let the data generating process:

$$y_t = \alpha y_{t-1} + u_t, \tag{1}$$

where $|\alpha| < 1$ and $\{u_t\}$ is a standard Gaussian White Noise:

$$u_t \sim \mathsf{NID}(0,1).$$

Under each of the following hypotheses, determine whether the process y_t is weakly stationary:

(i) the process $\{y_t\}$ started at $t_0 = -\infty,$ so that at each point in time, the process has infinite history.

(ii) the process starts at $t_0 = 0,$ with fixed initial value $y_0 = \overline{y}_0$ and $\{y_t\}$ obeys (1) for all $t > 0.$

(iii) the process starts at $t = 0$ with $y_0 \sim \mathsf{N}\left(0, \frac{1}{1-\alpha^2}\right)$ and obeys (1) for all $t > 0.$

Answer. First, notice that y_t can be re-written as

$$y_t = \alpha^{t-t_0} y_{t_0} + \sum_{i=0}^{t-t_0-1} \alpha^i u_{t-i}$$

where t_0 is the origin of the process. Therefore, since $\mathsf{E}[u_i] = 0$ for all i

$$\begin{aligned} \mathsf{E}[y_t] &= \alpha^{t-t_0}\mathsf{E}[y_{t_0}] + \sum_{i=0}^{t-t_0-1} \alpha^i \mathsf{E}[u_{t-i}] \\ &= \alpha^{t-t_0}\mathsf{E}[y_{t_0}] \end{aligned}$$

and

$$\mathsf{V}[y_t] = \alpha^{2(t-t_0)}\mathsf{V}[y_{t_0}] + \mathsf{V}\left[\sum_{i=0}^{t-t_0-1} \alpha^i u_{t-i}\right] + 2\alpha^{t-t_0}\sum_{i=0}^{t-t_0-1} \alpha^i \mathsf{Cov}[y_{t_0}, u_{t-i}].$$

Since $\{u_t\}$ is $i.i.d$, therefore $\mathsf{Cov}[y_{t_0}, u_i] = 0$ for $i > t_0$ and $\mathsf{Cov}[u_j, u_i] = 0$ if $i \neq j$. Then

$$\begin{aligned} \mathsf{V}[y_t] &= \alpha^{2(t-t_0)}\mathsf{V}[y_{t_0}] + \sum_{i=0}^{t-t_0-1} \alpha^{2i}\mathsf{V}[u_{t-i}] \\ &= \alpha^{2(t-t_0)}\mathsf{V}[y_{t_0}] + \frac{1-\alpha^{2(t-t_0)}}{1-\alpha^2}. \end{aligned}$$

This allows us to answer:

(i) here $t_0 \to -\infty$ and $y_t \sim \mathsf{N}\left(0, \frac{1}{1-\alpha^2}\right)$ is weakly stationary

(ii) here $t_0 = 0$ hence $y_t \sim \mathsf{N}\left(\alpha^t \overline{y}_0, \frac{1-\alpha^{2t}}{1-\alpha^2}\right)$ does not possess a constant distribution.

(iii) Now $y_t \sim \mathsf{N}\left(0, \frac{\alpha^{2t}}{1-\alpha^2} + \frac{1-\alpha^{2t}}{1-\alpha^2}\right) = \mathsf{N}\left(0, \frac{1}{1-\alpha^2}\right)$ hence if the origin has the asymptotic (stationary) distribution, then the process is weakly stationary. ∎

EXERCISE 3

Let $\{u_t\}$ a Gaussian white noise, i.e. $u_t \sim \mathsf{NID}(0, \sigma^2)$. The stochastic process $\{y_t\}$ derived from $\{u_t\}$ is defined as:

(i) $y_t = \alpha y_{t-1} + u_t,$ for $t > 0$, where $\alpha = \frac{1}{2}$ and $y_0 = 1$.

(ii) $y_t = u_t - \beta u_{t-1},$ with $\beta = 1$ for $t = 0, \pm 1, \pm 2, ...$

(iii) $y_t = \begin{cases} 1 + u_t & \text{for } t = 1, 3, 5, ... \\ -1 + u_t & \text{for } t = 2, 4, 6, ... \end{cases}$

For each of the processes defined above:

(a) Find the mean μ_t and autocovariance function $\gamma_t(h)$ of $\{y_t\}$;

(b) Determine whether the process is (weakly) stationary.

Answer. (i) y_t follows an AR(1) process. We have seen this in lecture and in the previous exercise.

$$\begin{aligned} \mu_t &= (1/2)^t \\ \gamma_t(0) &= \frac{1-(1/2)^{2t}}{1-(1/2)^2} = \frac{4}{3}\left\{1-(1/2)^{2t}\right\} \\ \gamma_t(1) &= \mathsf{Cov}[y_t, y_{t-1}] = \mathsf{Cov}[\alpha y_{t-1} + u_t, y_{t-1}] = \alpha\gamma_{t-1}(0) \\ \gamma_t(h) &= \alpha^h \gamma_{t-h}(0) \end{aligned}$$

(ii) y_t follows an MA(1) process with infinite memory. This is specific since the root is -1, hence the process is non invertible.

$$y_t = u_t - u_{t-1}$$

We know that an MA is stationary, hence

$$\begin{aligned}
\mu_t &= 0 \\
\gamma_t(0) &= \mathsf{Var}\left[u_t - u_{t-1}\right] = 2\sigma_u^2 \\
\gamma_t(1) &= \mathsf{Cov}\left[u_t - u_{t-1}, u_{t-1} - u_{t-2}\right] \\
&= -\mathsf{Var}\left[u_{t-1}\right] = -\sigma_u^2. \\
\gamma_t(h) &= 0 \quad for\ h \geq 2.
\end{aligned}$$

(iii) This process is more atypical

$$y_t = \begin{cases} 1 + u_t & for\ t\ odd \\ -1 + u_t & for\ t\ even \end{cases}$$

hence

$$\mu_t = y_t = \begin{cases} 1 & for\ t\ odd \\ -1 & for\ t\ even \end{cases}$$

$$\gamma_t(0) = \sigma_u^2$$

and

$$\gamma_t(h) = 0 \quad for\ h \geq 1$$

The expectation of the process depends on t, therefore the process is nonstationary. ■

EXERCISE 4

Consider the following ARMA$(2,1)$ process:

$$y_t = \alpha_2 y_{t-2} + \varepsilon_t + \beta_1 \varepsilon_{t-1},$$

where it is assumed that (i) $\beta_1 \neq \pm\sqrt{\alpha_2}$ and that (ii) $\{\varepsilon_t\}$ is a standard Gaussian white noise:

$$\varepsilon_t \sim \mathsf{NID}(0,1).$$

(a) Why is it assumed that (i) $\beta_1 \neq \pm\sqrt{\alpha_2}$?

(b) Under what conditions is the process $\{y_t\}$ stationary?

(c) Assuming that $\{y_t\}$ is stationary, what is the autocovariance function $\gamma(h)$ for $h \geq 0$.

Answer. (a) Re-write the ARMA$(2,1)$ process in terms of the lag polynomials:

$$\left(1 - \alpha_2 L^2\right) y_t = (1 - \beta_1 L)\, \varepsilon_{t-1}$$

for this to be truly an ARMA$(2,1)$, the AR and MA polynomials must have no (possibly complex) common roots, since these are $\left(-1/\sqrt{\alpha_2}, +1/\sqrt{\alpha_2}\right)$ and $1/\beta_1$, the assumption $\beta_1 \neq \pm\sqrt{\alpha_2}$ ensures that the model corresponds to the dynamics.

(b) For y_t to be stationary, the roots of the AR polynomial must be strict greater than unity in modulus. Hence $|\alpha_2| < 1$.

(c) Now, we first note that $\mathsf{E}[y] = 0$ under the assumption of stationarity. Now

$$\mathsf{V}[y_t] = \alpha_2^2 \mathsf{V}[y_{t-2}] + \mathsf{V}[\varepsilon_t] + \beta_1^2 \mathsf{V}[\varepsilon_{t-1}] + 2\alpha_2 \mathsf{Cov}[y_{t-2}, \varepsilon_t] + 2\alpha_2\beta_1 \mathsf{Cov}[y_{t-2}, \varepsilon_{t-1}] + 2\beta_1 \mathsf{Cov}[\varepsilon_t, \varepsilon_{t-1}]$$

hence

$$\begin{aligned} \gamma(0) &= \alpha_2^2 \gamma(0) + 1 + \beta_1^2 \\ \gamma(0) &= \frac{1 + \beta_1^2}{1 - \alpha_2^2} \end{aligned}$$

Now

$$\gamma(1) = \mathsf{Cov}[y_t, y_{t-1}] = \alpha_2 \gamma(1) + \mathsf{Cov}[\varepsilon_t, y_{t-1}] + \beta_1 \mathsf{Cov}[\varepsilon_{t-1}, y_{t-1}]$$

where we know that $\mathsf{Cov}[\varepsilon_{t-1}, y_{t-1}] = \mathsf{Cov}[\varepsilon_t, y_t] = \mathsf{Cov}[\varepsilon_t, \alpha_2 y_{t-2} + \varepsilon_t + \beta_1 \varepsilon_{t-1}] = \sigma_\varepsilon^2 = 1$, hence

$$\gamma(1) = \frac{\beta_1}{1 - \alpha_2}.$$

We continue:

$$\begin{aligned} \gamma(2) &= \mathsf{Cov}[y_t, y_{t-2}] = \alpha_2 \gamma(0) \\ \gamma(3) &= \alpha_2 \gamma(1) \\ \gamma(h) &= \alpha_2 \gamma(h-2), \text{ for } h \geq 3 \end{aligned}$$

■

EXERCISE 5

You will find below eight figures with four graphs each. These represent from left to right, top to bottom, (a) the time series, (b) its autocorrelation function (ACF), (c) its partial autocorrelation function, and (d) a scatter plot of y_t against y_{t-1}. These graphs were drawn from series generated using the same white noise series $\{\varepsilon_t\}$ as:

$$\begin{aligned} &(1) : y_t = y_{t-1} + \varepsilon_t - 0.5\varepsilon_{t-1} \\ &(2) : y_t = 0.95 y_{t-1} + \varepsilon_t \\ &(3) : y_t = 0.6 y_{t-1} + \varepsilon_t \\ &(4) : y_t = 0.6 y_{t-4} + \varepsilon_t - 0.3\varepsilon_{t-1} \\ &(5) : y_t = 0.2 y_{t-1} + 0.8 y_{t-4} + \varepsilon_t \\ &(6) : y_t = 1.3 y_{t-1} - 0.4 y_{t-2} + \varepsilon_t - 0.5\varepsilon_{t-1} \\ &(7) : y_t = \varepsilon_t + 0.8\varepsilon_{t-1} \\ &(8) : y_t = y_{t-1} + \varepsilon_t \end{aligned}$$

(a) Using the information at your disposal, including any computation that you may think useful, find which of series 1 to 8 corresponds to graphs A to H.

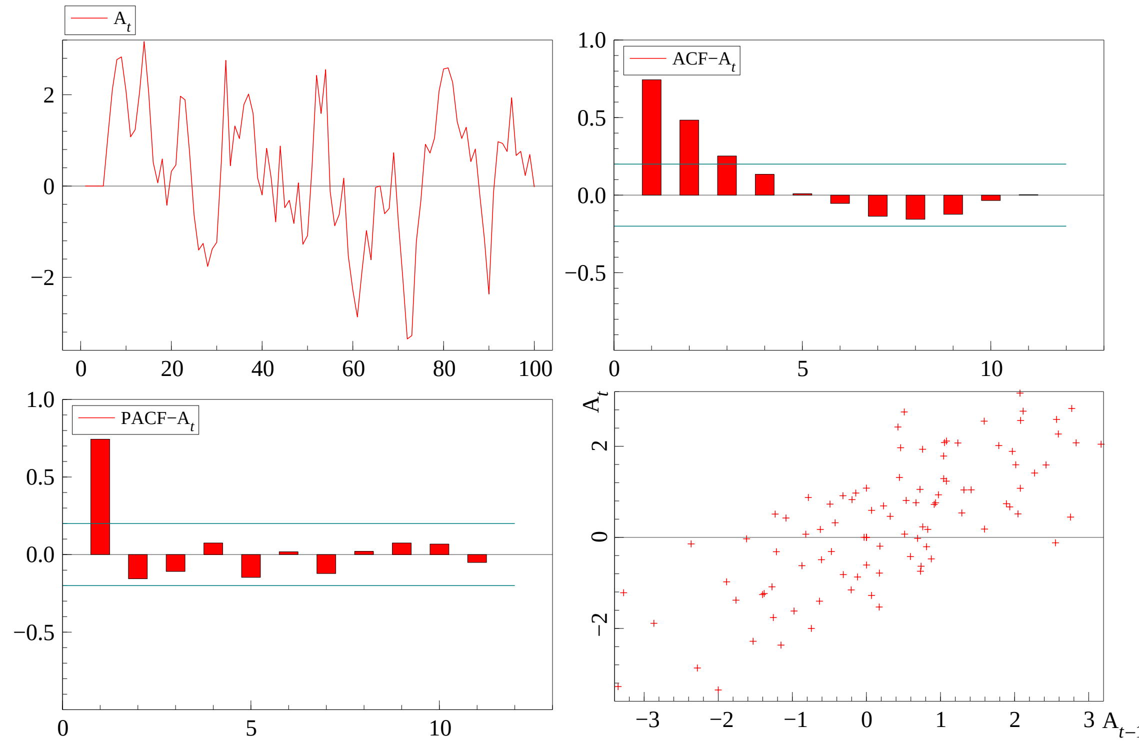


Figure 1: Series A

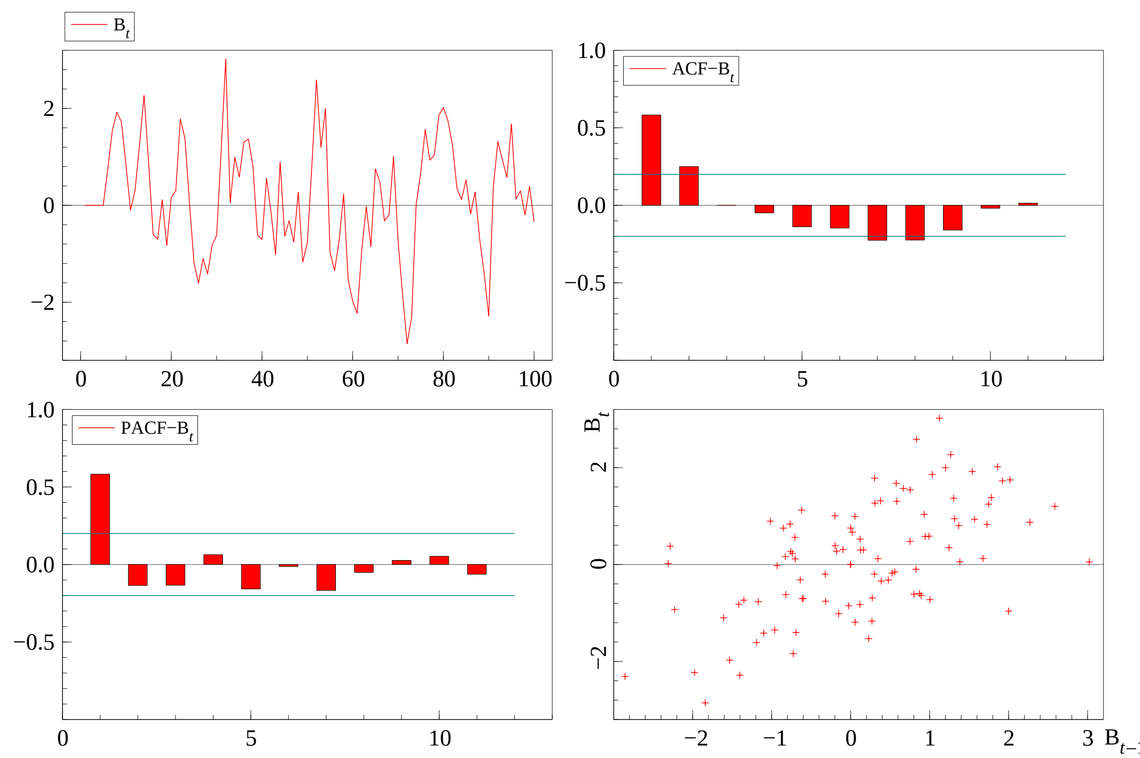


Figure 2: Series B

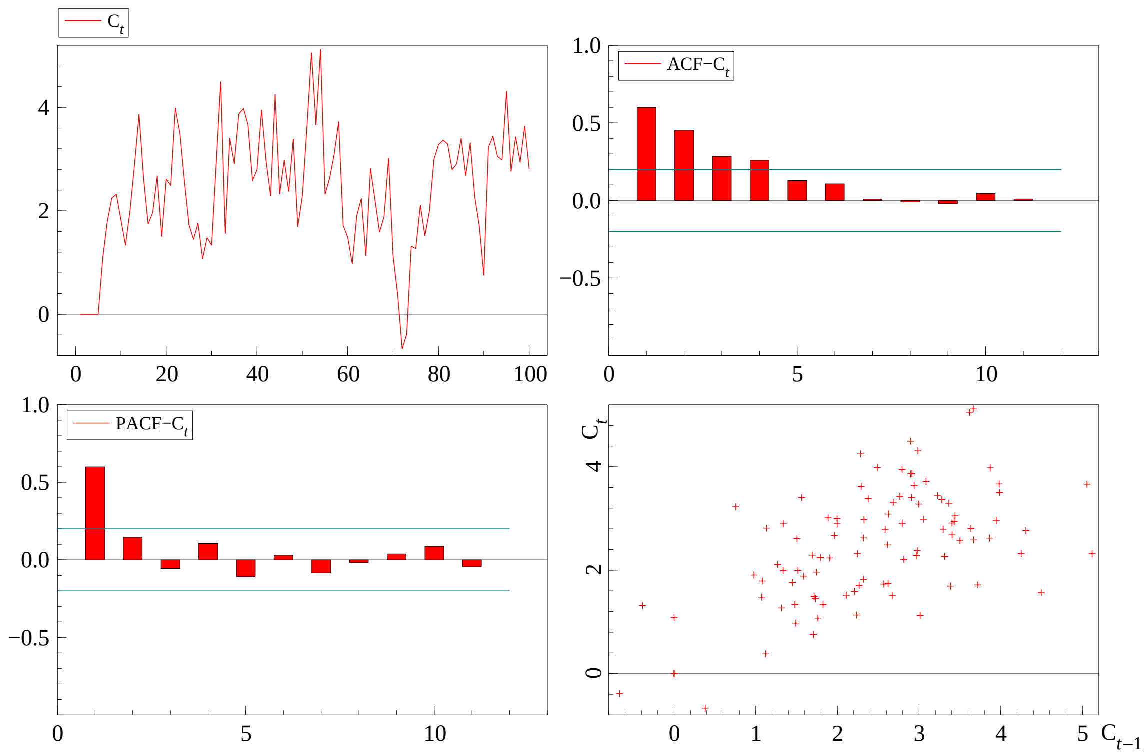


Figure 3: Series C

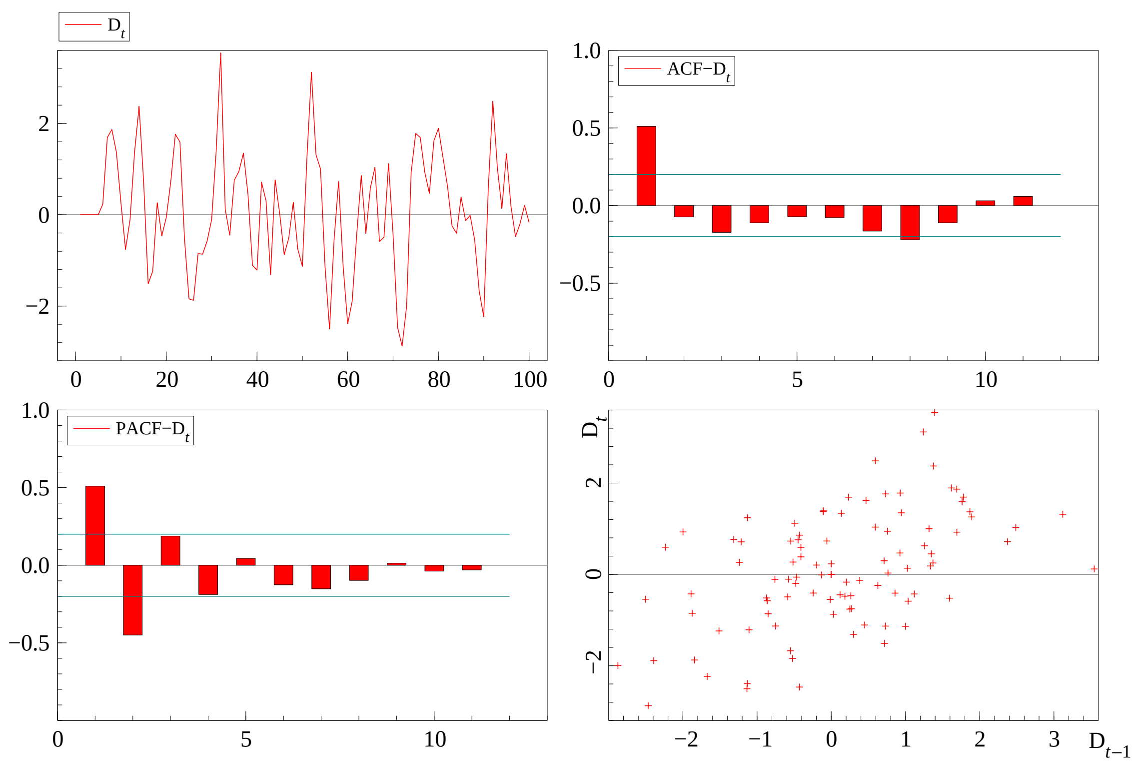


Figure 4: Series D

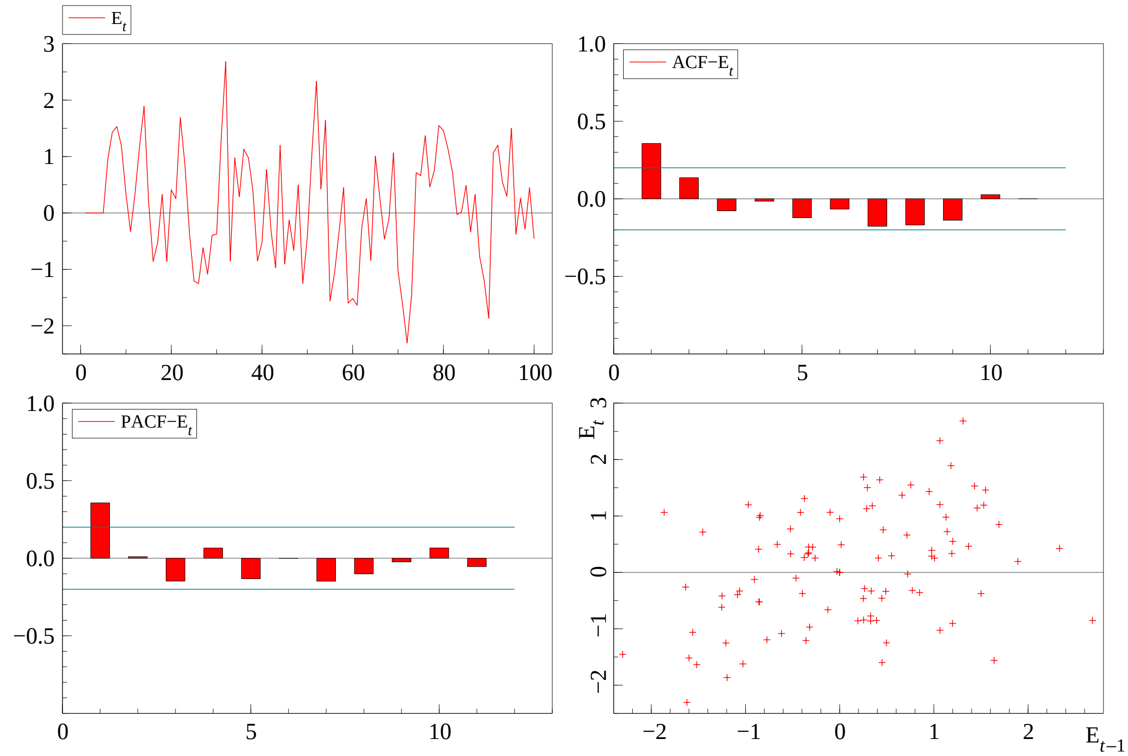


Figure 5: Series E

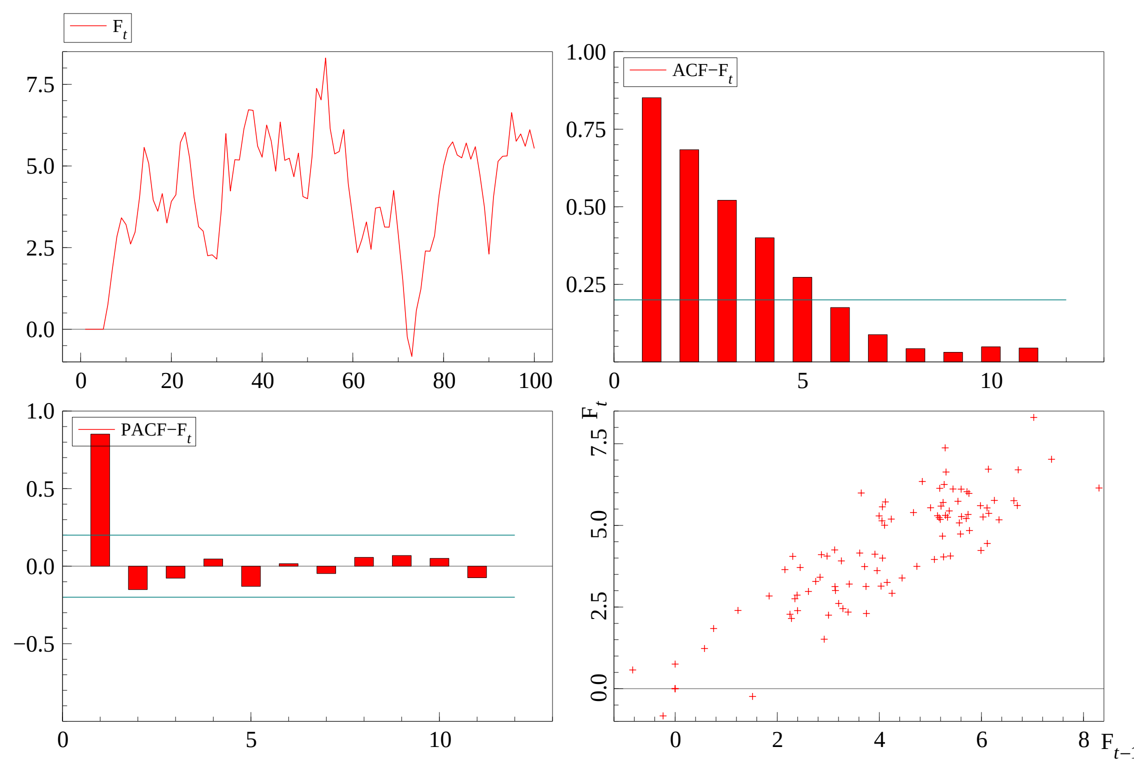


Figure 6: Series F

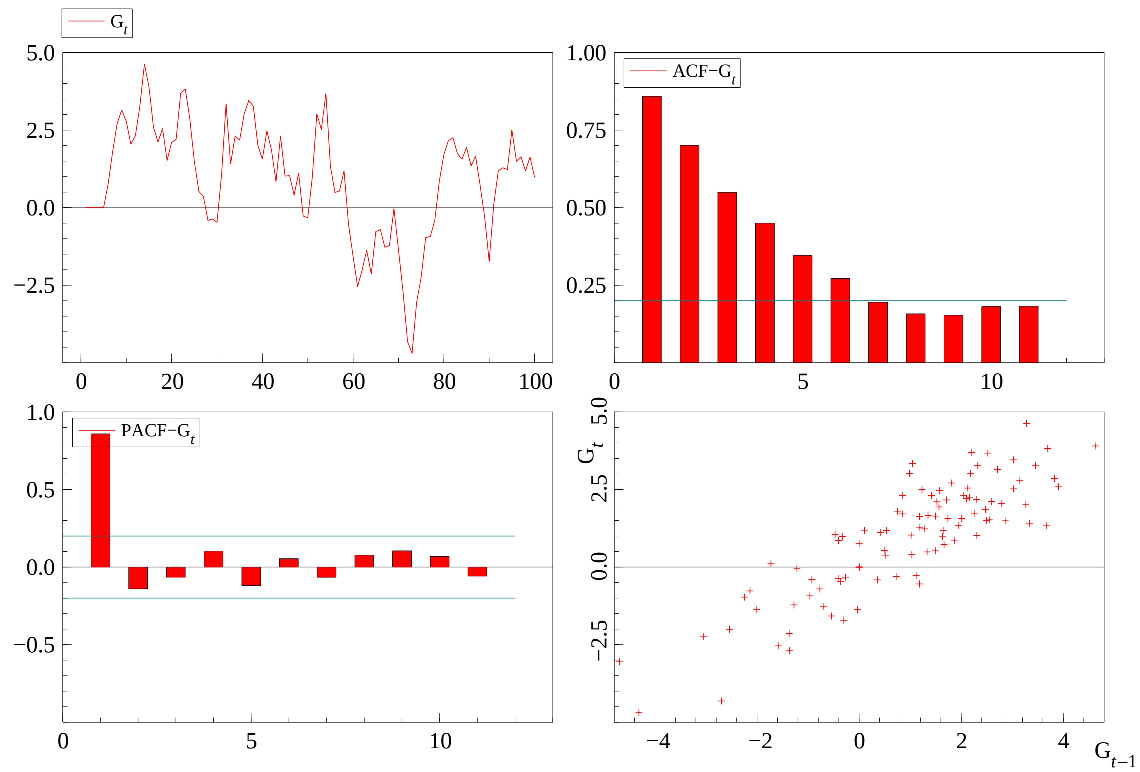


Figure 7: Series G

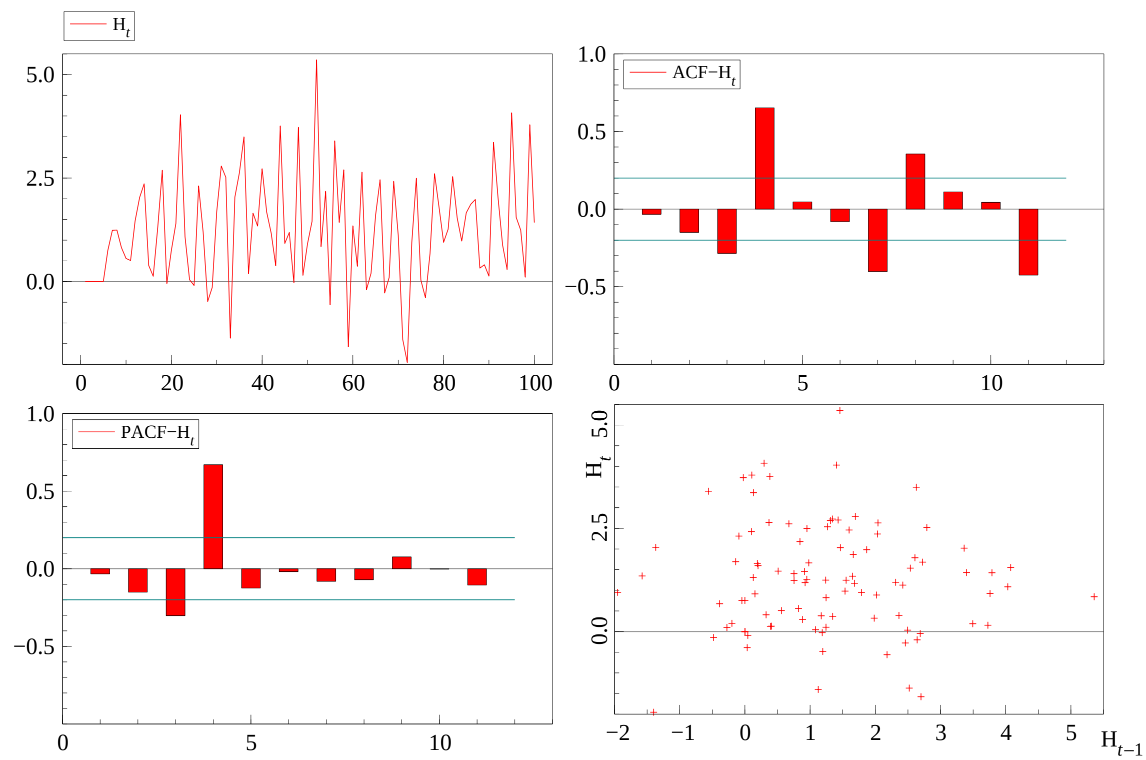


Figure 8: Series H

(b) Which of the series are stationary?

(c) Which of the series present empirical properties that differ from their theoretical properties?

Answer. We start by observing the series: ■

A- the series seems stationary, ACF seems to imply an AR, an AR(1) according to PACF with a coefficient of about 0.75.

B- the series seems stationary, AR(1) with a coefficient of about 0.6.

C- the series does not have zero mean; it is not clear that it is stationary, yet ACF/PACF imply either an AR(1) with a parameter of 0.6 or an MA(4).

D- the series is clearly stationary, ACF would mean an MA(1) , PACF too.

E- the series is clearly stationary, very little memory via either ACF or PACF.

F- the series is non-stationary, it looks like a random walk.

G- as series F but ACF shows more inertia. Yet the series oscillates around zero.

H- the series is roughly stationary (yet not around zero); strong ACF at lags 3,4,7,8,11... and PACF at 3,4...

Now, regarding the ARMA:

(1) ARMA$(1,1)$ with unit root

(2) AR(1) nearly integrated

(3) AR(1) with parameter 0.6 (series B?)

(4) ARMA(4,1)

(5) AR(4), the sum of lag coefficients is unity (unit root)

(6) ARMA(2,1) with large AR coefficients, yet their sum is less than unity

(7) MA(1) (series D?)

(8) random walk (series F?)

Now, two series (other than F) have nonzero mean: C and H, whereas according to the definitions, they should all have zero expectations. Hence we can assume that these two series are non-stationary. The other two unit-roots are (1) and (5) . C seems to correspond to (1) , H would then be (5) .

There remains A, E, G and (2) , (4) , (6) . (2) is close to a random walk but oscillates around zero (hence stationary in this respect), could it be G?

For (6) , we factorize the polynomials:

$$\begin{aligned}
\left(1-1.3L+.4L^2\right) y_t &= (1-0.5L)\,\varepsilon_t \\
(1-0.5L)(1-0.8L)\, y_t &= (1-0.5L)\,\varepsilon_t \\
(1-0.8L)\, y_t &= \varepsilon_t
\end{aligned}$$

Hence, this corresponds to an AR(1) with parameter 0.8: series A. There remains only E for (4) , here the two AR and MA effects seem to cancel.

Conclusion: (1) C; (2) G; (3) B; (4) E; (5) H; (6) A; (7) D; (8) F.

EXERCISE 6

From an observed series $\{y_t\}$, two stationary processes are considered as potential candidates for the data generating process (DGP):

$$y_t = \nu + \alpha y_{t-2} + u_t, \quad u_t \sim \mathsf{NID}\left(0, \sigma_u^2\right), \tag{2}$$

$$y_t = \mu + \varepsilon_t + \beta\varepsilon_{t-2}, \quad \varepsilon_t \sim \mathsf{NID}\left(0, \sigma_\varepsilon^2\right). \tag{3}$$

Let the empirical moments be, for three different cases:

$$\begin{aligned}
(i) \quad & \overline{y} = 0, \quad \widehat{\mathsf{Var}}\left[y_t\right] = 2, \quad \widehat{\mathsf{Corr}}\left[y_t, y_{t-1}\right] = 0,9, \quad \widehat{\mathsf{Corr}}\left[y_t, y_{t-2}\right] = 0,4\ ; \\
(ii) \quad & \overline{y} = 1, \quad \widehat{\mathsf{Var}}\left[y_t\right] = 2, \quad \widehat{\mathsf{Corr}}\left[y_t, y_{t-1}\right] \approx 0, \quad \widehat{\mathsf{Corr}}\left[y_t, y_{t-2}\right] = 0,4\ ; \\
(iii) \quad & \overline{y} = 1, \quad \widehat{\mathsf{Var}}\left[y_t\right] = 2, \quad \widehat{\mathsf{Corr}}\left[y_t, y_{t-1}\right] \approx 0, \quad \widehat{\mathsf{Corr}}\left[y_t, y_{t-2}\right] = 0,8\ ;
\end{aligned}$$

Find, if possible, the values the model parameters (1) , $\left\{\nu, \alpha, \sigma_u^2\right\}$ and (2) , $\left\{\mu, \beta, \sigma_\varepsilon^2\right\}$ from the empirical moments and discuss the additional information that could help you identify the DGP.

Answer. (1) We derive the moments of the series generated by the AR(2) :

$$y_t = \nu + \alpha y_{t-2} + u_t, \quad u_t \sim \mathsf{NID}\left(0, \sigma_u^2\right).$$

The process is assumed stationary. ■

(i) Expectation:

$$\mathsf{E}\left[y_t\right] = \mathsf{E}\left[y_{t-2}\right] = \frac{\nu}{1-\alpha}$$

(ii) Variance:

$$\mathsf{V}\left[y_t\right] = \mathsf{V}\left[y_{t-2}\right] = \frac{\sigma_u^2}{1-\alpha^2}$$

(iii) Covariance between y_t and y_{t-1} : $\mathsf{Cov}\left[y_t, y_{t-1}\right] = 0$ as y_t does not depend on y_{t-1} and u_t is White Noise. (easy to show by induction).

(iii) Covariance between y_t and y_{t-2} :

$$\begin{aligned}
\mathsf{Cov}\left[y_t, y_{t-2}\right] &= \mathsf{Cov}\left[y_t - \frac{\nu}{1-\alpha}, y_{t-2} - \frac{\nu}{1-\alpha}\right] \\
&= \mathsf{E}\left[\left(y_t - \frac{\nu}{1-\alpha}\right)\left(y_{t-2} - \frac{\nu}{1-\alpha}\right)\right] \\
&= \mathsf{E}\left[\left(y_t - \frac{\nu}{1-\alpha}\right)\left(y_{t-2} - \frac{\nu}{1-\alpha}\right)\right],
\end{aligned}$$

but

$$y_t - \frac{\nu}{1-\alpha} = \alpha\left(y_{t-2} - \frac{\nu}{1-\alpha}\right) + u_t,$$

hence, as y_{t-2} and u_t are independent:

$$\begin{aligned}
\mathsf{Cov}\left[y_t, y_{t-2}\right] &= \alpha \mathsf{V}\left[y_{t-2} - \frac{\nu}{1-\alpha}\right] \\
&= \frac{\alpha}{1-\alpha^2}\sigma_u^2
\end{aligned}$$

(2) We derive the moments of the series generated by the $\mathrm{MA}(2)$:

$$y_t = \mu + \varepsilon_t + \beta\varepsilon_{t-2}, \quad \varepsilon_t \sim \mathsf{NID}\left(0, \sigma_\varepsilon^2\right).$$

The process is stationary since it is a moving average.

(i) Expectation:

$$\mathsf{E}\left[y_t\right] = \mu$$

(ii) Variance:

$$\mathsf{V}\left[y_t\right] = \left(1 + \beta^2\right)\sigma_\varepsilon^2$$

(iii) Covariance between y_t and y_{t-1} : $\mathsf{Cov}\left[y_t, y_{t-1}\right] = 0$ as y_t and y_{t-2} are function of independent White Noises.

(iii) Covariance between y_t and y_{t-2} :

$$\begin{aligned}\mathsf{Cov}\left[y_t, y_{t-2}\right] &= \mathsf{Cov}\left[\varepsilon_t + \beta\varepsilon_{t-2}, \varepsilon_{t-2} + \beta\varepsilon_{t-4}\right] \\ &= \beta\sigma_\varepsilon^2.\end{aligned}$$

Now, to compute the parameters from the moments, we need to solve the equations. For this we use the ratio $\frac{\mathsf{Cov}[y_t, y_{t-2}]}{\mathsf{V}[y_t]}$ and compute the coefficients (second order equation in the case of the $\mathrm{MA}(2)$).

$$AR\left(2\right): \quad \begin{cases} \nu = \mathsf{E}\left[y_t\right]\left(1 - \mathsf{Corr}\left[y_t, y_{t-2}\right]\right) \\ \\ \alpha = \mathsf{Corr}\left[y_t, y_{t-2}\right] \\ \\ \sigma_u^2 = \mathsf{V}\left[y_t\right]\left(1 - \left\{\mathsf{Corr}\left[y_t, y_{t-2}\right]\right\}^2\right) \end{cases}$$

$$MA\left(2\right) \quad \begin{cases} \mu = \mathsf{E}\left[y_t\right] \\ \\ \beta = \dfrac{1 \pm \sqrt{1 - 4\left\{\mathsf{Corr}\left[y_t, y_{t-2}\right]\right\}^2}}{2\mathsf{Corr}\left[y_t, y_{t-2}\right]} \\ \\ \sigma_\varepsilon^2 = \dfrac{2\left\{\mathsf{Corr}\left[y_t, y_{t-2}\right]\right\}^2}{1 \pm \sqrt{1 - 4\left\{\mathsf{Corr}\left[y_t, y_{t-2}\right]\right\}^2}} \end{cases}$$

To get β, two solutions exist, one greater than unity, the other smaller; we use the latter to ensure invertibility of the MA.

Empirical applications

(i) We notice that the estimated first-order autocorrelation is large, whereas each model implies a population value of zero. Neither model can hence apply.

(ii) We choose, for the MA(2) , the coefficient that implies an invertible process.

$$AR(2): \begin{cases} \nu = 0,6 \\ \alpha = 0,4 \\ \sigma_u^2 = 1,68 \end{cases}$$

$$MA(2): \begin{cases} \mu = 1 \\ \beta = 0,5 \text{ (or 2)} \\ \sigma_\varepsilon^2 = 1,6 \text{ (or } 0,4) \end{cases}$$

To choose between the two representations, the correlogram would be useful.

(iii) Only the AR representation is here possible.

$$AR(2): \begin{cases} \nu = 0,2 \\ \alpha = 0,8 \\ \sigma_u^2 = 0,72 \end{cases}$$

$$MA(2): \text{no real root for } \beta$$

EXERCISE 7

We define the exponential smoother (or Exponentially Weighted Moving Average (EWMA) model) for a series y_t as

$$\begin{aligned} \widetilde{y}_{T+1} &= \alpha y_T + \alpha(1-\alpha) y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} \ldots. \\ &= \alpha \sum_{i=0}^{\infty} (1-\alpha)^i y_{T-i}, \end{aligned}$$

for $\alpha \in (0,1)$.

1. Show that the exponential smoother is mean preserving.
2. What is the effect of varying α?
3. At a date T, the exponential smoother is used for forecasting y_{T+h} for $h > 0$. Show that the forecast is the same at all horizons.

Answer.

1. We compute the expectation $\mathsf{E}[\widetilde{y}_{T+1}] = \alpha \sum_{i=0}^{\infty} (1-\alpha)^i \mathsf{E}[y_{T-i}]$. If the series y_t is stationary with expectation μ, then
$$\mathsf{E}[\widetilde{y}_{T+1}] = \mu\alpha \frac{1}{1-(1-\alpha)} = \mu.$$
2. By lowering alpha, we see that the weights attached to past values is then larger. A lower α therefore means that the exponential smoother takes more into account the history of y_t. At the limit when $\alpha \to 0$, we see that $\widetilde{y}_{T+1}$ is almost the average of all past values, which does not oscillate much when we add or remove one observation. By contrast when $\alpha \to 1$, $\widetilde{y}_{T+1} \to y_T$. Hence, the exponential smoother is a time series that has the same expectation as the original time series (y_t) but which oscillates less, at one limit it is very close to the sample mean of y_t, and at the other extreme, it is no different from y (lagged by one value: $\widetilde{y}_{T+1} = y_T$).

3. Let $h \geq 1$ and forecast y_{T+h} using

$$\widehat{y}_{T+h|T} = \alpha \sum_{i=1}^{h-1} (1-\alpha)^{i-1} \widehat{y}_{T+h-i}$$

where, between $T+1$ and $T+h-1$, we use the forecast $\widehat{y}_{T+h-i}$ and for $t \leq T$ $\widehat{y}_t = y_t$. Therefore

$$\begin{aligned}
\widehat{y}_{T+h|T} &= \alpha \sum_{i=1}^{h-1} (1-\alpha)^{i-1} \widehat{y}_{T+h-i} + \alpha \sum_{i=h}^{\infty} (1-\alpha)^{i-1} y_{T+h-i} \\
&= \alpha \sum_{i=1}^{h-1} (1-\alpha)^{i-1} \left[\alpha \sum_{j=0}^{\infty} (1-\alpha)^j y_{T-j} \right] + \alpha \sum_{i=0}^{\infty} (1-\alpha)^{h-1+i} y_{T-i} \\
&= \left[\alpha \sum_{i=1}^{h-1} (1-\alpha)^{i-1} + (1-\alpha)^{h-1} \right] \left[\alpha \sum_{j=0}^{\infty} (1-\alpha)^j y_{T-j} \right] \\
&= \left[\alpha \frac{1-(1-\alpha)^{h-1}}{1-(1-\alpha)} + (1-\alpha)^{h-1} \right] \left[\alpha \sum_{j=0}^{\infty} (1-\alpha)^j y_{T-j} \right] \\
&= \left[1-(1-\alpha)^{h-1} + (1-\alpha)^{h-1} \right] \left[\alpha \sum_{j=0}^{\infty} (1-\alpha)^j y_{T-j} \right] \\
&= \left[\alpha \sum_{j=0}^{\infty} (1-\alpha)^j y_{T-j} \right] \\
&= \widetilde{y}_{T+1}
\end{aligned}$$

Hence, for all $h \geq 1$

$$\widehat{y}_{T+h|T} = \widetilde{y}_{T+1}.$$

■

EXERCISE 8

Let $\{u_t\}$ a Gaussian white noise, i.e. $u_t \sim \mathsf{NID}\left(0, \sigma^2\right)$. The stochastic process $\{y_t\}$ derived from $\{u_t\}$ is defined as:

$$\begin{array}{llll}
(i) & y_t = & \alpha y_{t-1} + u_t, & \text{for } t > 0 \text{, where } \alpha = 1 \text{ and } y_0 = 2. \\
(ii) & y_t = & \tau + \alpha y_{t-1} + u_t, & \text{for } t > 0 \text{, where } \alpha = 1, \tau \neq 0 \text{, and } y_0 = 0\,;
\end{array}$$

For each of the processes defined above:

(a) Find the mean μ_t and autocovariance function $\gamma_t(h)$ of $\{y_t\}$;

(b) Determine whether the process is (weakly) stationary.

Answer. (i)

$$y_t = y_{t-1} + u_t, \quad with\; y_0 = 2$$

hence

$$y_t = y_0 + \sum_{i=1}^{t} u_i$$

and

$$
\begin{aligned}
\mu_t &= \mathsf{E}[y_t] = y_0 = 2 \\
\gamma_t(0) &= \mathsf{Var}\left[\sum_{i=1}^{t} u_i\right] = \sum_{i=1}^{t} \mathsf{Var}[u_i] = t\sigma_u^2 \\
\gamma_t(h) &= \mathsf{Cov}[y_t, y_{t-h}] = \mathsf{Cov}\left[\sum_{i=1}^{t} u_i, \sum_{i=1}^{t-h} u_i\right] \\
&= \mathsf{Cov}\left[\sum_{i=1}^{t-h} u_i, \sum_{i=1}^{t-h} u_i\right] + \mathsf{Cov}\left[\sum_{i=t-h+1}^{t} u_i, \sum_{i=1}^{t-h} u_i\right]
\end{aligned}
$$

but the $(u_{t-h+1}, ..., u_t)$ are independent from $(u_1, ..., u_{t-h})$ hence

$$
\begin{aligned}
\gamma_t(h) &= \mathsf{Cov}\left[\sum_{i=1}^{t-h} u_i, \sum_{i=1}^{t-h} u_i\right] = \mathsf{Var}\left[\sum_{i=1}^{t-h} u_i\right] = (t-h)\,\sigma_u^2 \quad pour\ h \geq 0 \\
\gamma_t(h) &= t\sigma_u^2 \quad pour\ h < 0.
\end{aligned}
$$

and the process is non-stationary: it is a random walk.

(ii) Here, the process is a random walk with drift (non-stationary)

$$
\begin{aligned}
y_t &= \tau + y_{t-1} + u_t \\
&= \tau t + \sum_{i=1}^{t} u_i
\end{aligned}
$$

hence

$$
\begin{aligned}
\mu_t &= \mathsf{E}[y_t] = \tau t \\
\gamma_t(0) &= \mathsf{Var}\left[\sum_{i=1}^{t} u_i\right] = \sum_{i=1}^{t} \mathsf{Var}[u_i] = t\sigma_u^2 \\
\gamma_t(h) &= \mathsf{E}\left[\left(\tau t + \sum_{i=1}^{t} u_i\right)\left(\tau(t-h) + \sum_{i=1}^{t-h} u_i\right)\right] - \tau t\,[\tau(t-h)] \\
&= (t-h)\,\sigma_u^2 \quad for\ h \geq 0 \quad and \quad t\sigma_u^2 \quad for\ \ h < 0.
\end{aligned}
$$

■

EXERCISE 10

We dispose of a series $\{y_t\}$ for $t = -\infty$ to T, but only consider the observations for $t = 1, ..., T$. This series seems to present a linear deterministic trend but we hesitate between two models for its representation:

$$\mathsf{TS} : y_t = \alpha + \beta t + u_t, \tag{4}$$

$$\mathsf{DS} : y_t = y_{t-1} + u_t \tag{5}$$

where $u_t \sim \mathsf{IN}\left(0, \sigma_u^2\right)$.

1. *Briefly* comment on the properties of the two TS and DS models. Why do we hesitate between them?

2. What would your approach be, should you wish to find which models fit the series best?

   In order to make a choice between the models, we focus on their forecasting properties. We assume that, at time T, we wish to generate a forecast h periods into the future, with $h > 0$.

3. If y_t follows model TS, what is the value of y_{T+h} as a function of y_T? Same question if y_t follows model DS.

   We assume that for each model, we generate a forecast for y_{T+h}. We denote them by $\widehat{y}^{\mathsf{TS}}_{T+h}$ and $\widehat{y}^{\mathsf{DS}}_{T+h}$ respectively. In order to compute them, we assume that the forecast is the expectation of y_{T+h} conditional on y_T; more precisely:

$$\widehat{y}^{\mathsf{TS}}_{T+h} = \mathsf{E}\left[y_{T+h}|y_T\right] \text{ assuming that } y_t \text{ follows model } \mathsf{TS},$$
$$\widehat{y}^{\mathsf{DS}}_{T+h} = \mathsf{E}\left[y_{T+h}|y_T\right] \text{ assuming that } y_t \text{ follows model } \mathsf{DS},$$

   Assume to simplify that $\mathrm{E}[u_T|y_T] = 0$.

4. Compute $\widehat{y}^{\mathsf{TS}}_{T+h}$ and $\widehat{y}^{\mathsf{DS}}_{T+h}$.

5. We assume in this question that y_t follows model TS

   (a) Using model TS to compute the forecast, we define the forecast error as:

$$e_{\mathsf{TS|TS}} = y_{T+h} - \widehat{y}^{\mathsf{TS}}_{T+h}$$

   Derive $e_{\mathsf{TS|TS}}$, its expectation and variance.

   (b) Alternatively, we erroneously use model DS to compute the forecast and define the forecast error, when y_{T+h} follows model TS:

$$e_{\mathsf{DS|TS}} = y_{T+h} - \widehat{y}^{\mathsf{DS}}_{T+h}$$

   Derive $e_{\mathsf{DS|TS}}$, its expectation and variance.

6. Now, we assume that y_t follows model DS and define as before:

$$e_{\mathsf{TS|DS}} = y_{T+h} - \widehat{y}^{\mathsf{TS}}_{T+h}$$

   and

$$e_{\mathsf{DS|DS}} = y_{T+h} - \widehat{y}^{\mathsf{DS}}_{T+h}$$

   Derive $e_{\mathsf{TS|DS}}$ and $e_{\mathsf{DS|DS}}$, their expectations and variances.

7. Assuming that it is possible that we are mistaken in our model choice. To minimize the error risk, which model would you pick? What do you think of this criterion for model choice.

Answer

1. TS is stationary around a linear trend: its variance is constant, but not its expectation. DS follows a random walk and is integrated: its difference is stationary, its expectation is constant and zero, but its variance grows linearly in time. Neither model is stationary, the issue is how to model the non-stationarity.

2. We need to do a unit-root test, allowing for the presence of a deterministic trend.

3.
$$(TS): y_{T+2} = y_T + 2\beta + u_{T+2} - u_T$$
$$(DS): y_{T+2} = y_T + u_{T+2} + u_{T+1}$$

hence

$$(TS): y_{T+h} = y_T + h\beta + u_{T+h} - u_T$$
$$(DS): y_{T+h} = y_T + \sum_{i=1}^{h} u_{T+i}$$

4. We assume $\mathrm{E}[u_T|y_T] = 0$,
$$\widehat{y}^{\mathsf{TS}}_{T+h} = \mathsf{E}\left[y_T + h\beta + u_{T+h} - u_T|y_T\right] = y_T + h\beta,$$
$$\widehat{y}^{\mathsf{DS}}_{T+h} = \mathsf{E}\left[y_T + \sum_{i=1}^{h} u_{T+i}|y_T\right] = y_T,$$

5.
$$e_{\mathsf{TS|TS}} = u_{T+h} - u_T$$
$$e_{\mathsf{DS|TS}} = h\beta + u_{T+h} - u_T$$

hence

$$\mathsf{E}\left[e_{\mathsf{TS|TS}}\right] = 0$$
$$\mathsf{E}\left[e_{\mathsf{DS|TS}}\right] = h\beta$$

$$\mathsf{V}\left[e_{\mathsf{TS|TS}}\right] = 2\sigma_u^2$$
$$\mathsf{V}\left[e_{\mathsf{DS|TS}}\right] = 2\sigma_u^2$$

6.
$$e_{\mathsf{TS|DS}} = \sum_{i=1}^{h} u_{T+i} - h\beta$$
$$e_{\mathsf{DS|DS}} = \sum_{i=1}^{h} u_{T+i}$$

hence

$$\mathsf{E}\left[e_{\mathsf{TS|DS}}\right] = -h\beta$$
$$\mathsf{E}\left[e_{\mathsf{DS|DS}}\right] = 0$$

$$\mathsf{V}\left[e_{\mathsf{TS|DS}}\right] = h\sigma_u^2$$
$$\mathsf{V}\left[e_{\mathsf{DS|DS}}\right] = h\sigma_u^2$$

where β needs be estimated or assumed.

7. The problem here is to minimize the forecast error, allowing for the possibility that we do not pick the right model. Given that whichever model is true, both models result in the same forecast variance, what we should attempt is to minimize the forecast bias. These are $\pm h\beta$, so each model provides the same absolute bias when it is wrongly used.

   What we must remember though, is that we need to estimate the parameters and that will have an impact. When the true model is TS, but that we use DS, then $\mathsf{E}\left[e_{\mathsf{DS}|\mathsf{TS}}\right] = h\beta$, but in fact, when the converse holds: $\mathsf{E}\left[e_{\mathsf{TS}|\mathsf{DS}}\right] = -h\widehat{\beta}$ where $\widehat{\beta}$ is estimated (do you see why β is not estimated in the first case?). The question that remains is: is $\left|\widehat{\beta}\right|$ estimated when the true model is DS smaller than $|\beta|$? You can try by simulations, but this should hold in practice. For $\left|\widehat{\beta}\right| > |\beta|$ we would need DS to produce an explosive pattern, which is quite unlikely.

EXERCISE 11: VAR

Consider the VAR(1) given by

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} .1 \\ 1 \end{bmatrix} + \begin{bmatrix} .8 & 0 \\ .2 & .4 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

1. Is the model stationary, cointegrated, or with two independent unit roots?
   ANSWER: the model rewrites, using a lag-polynomial

$$\left(\mathbf{I}_2 - \begin{bmatrix} .8 & 0 \\ .2 & .4 \end{bmatrix} L\right) \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} .1 \\ 1 \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

   We consider the roots of the determinant of the AR matrix lag polynomial:

$$\begin{aligned} \left|\mathbf{I}_2 - \begin{bmatrix} .8 & 0 \\ .2 & .4 \end{bmatrix} z\right| &= \left|\begin{bmatrix} 1-.8z & 0 \\ -.2z & 1-.4z \end{bmatrix}\right| \\ &= (1-.8z)(1-.4z) \end{aligned}$$

   so the roots are $(z_1, z_2) = (5/4, 5/2)$ and both are greater than unity is absolute value (or modulus) so the vector process admits a stationary solution (i.e. with given assumptions about the initial (x_0, y_0)).

2. Trace out the 6 steps of the impulse response function for a unit shock to ϵ_{1t} and ϵ_{2t}
   ANSWER: Express (x_t, y_t) as functions of past errors

$$\begin{aligned} \begin{bmatrix} x_t \\ y_t \end{bmatrix} &= \begin{bmatrix} .1 \\ 1 \end{bmatrix} + \begin{bmatrix} .8 & 0 \\ .2 & .4 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \\ &= \left(\mathbf{I}_2 + \begin{bmatrix} .8 & 0 \\ .2 & .4 \end{bmatrix}\right) \begin{bmatrix} .1 \\ 1 \end{bmatrix} + \begin{bmatrix} .8 & 0 \\ .2 & .4 \end{bmatrix}^2 \begin{bmatrix} x_{t-2} \\ y_{t-2} \end{bmatrix} \\ &\quad + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} + \begin{bmatrix} .8 & 0 \\ .2 & .4 \end{bmatrix} \begin{bmatrix} \epsilon_{1t-1} \\ \epsilon_{2t-1} \end{bmatrix} \\ &= \begin{bmatrix} 0.18 \\ 1.42 \end{bmatrix} + \begin{bmatrix} 0.64 & 0 \\ 0.24 & 0.16 \end{bmatrix} \begin{bmatrix} x_{t-2} \\ y_{t-2} \end{bmatrix} \\ &\quad + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} + \begin{bmatrix} .8 & 0 \\ .2 & .4 \end{bmatrix} \begin{bmatrix} \epsilon_{1t-1} \\ \epsilon_{2t-1} \end{bmatrix} \end{aligned}$$

and also

$$
\begin{aligned}
\begin{bmatrix} x_t \\ y_t \end{bmatrix} &= \begin{bmatrix} 0.244 \\ 1.604 \end{bmatrix} + \begin{bmatrix} 0.512 & 0 \\ 0.224 & 0.064 \end{bmatrix} \begin{bmatrix} x_{t-3} \\ y_{t-3} \end{bmatrix} \\
&\quad + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} + \begin{bmatrix} .8 & 0 \\ .2 & .4 \end{bmatrix} \begin{bmatrix} \epsilon_{1t-1} \\ \epsilon_{2t-1} \end{bmatrix} + \begin{bmatrix} .8 & 0 \\ .2 & .4 \end{bmatrix}^2 \begin{bmatrix} \epsilon_{1t-2} \\ \epsilon_{2t-2} \end{bmatrix} \\
&= \begin{bmatrix} 0.295\,2 \\ 1.69 \end{bmatrix} + \begin{bmatrix} 0.409\,6 & 0 \\ 0.192 & 0.025\,6 \end{bmatrix} \begin{bmatrix} x_{t-4} \\ y_{t-4} \end{bmatrix} \\
&\quad + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} + \begin{bmatrix} .8 & 0 \\ .2 & .4 \end{bmatrix} \begin{bmatrix} \epsilon_{1t-1} \\ \epsilon_{2t-1} \end{bmatrix} + \begin{bmatrix} 0.64 & 0 \\ 0.24 & 0.16 \end{bmatrix} \begin{bmatrix} \epsilon_{1t-2} \\ \epsilon_{2t-2} \end{bmatrix} \\
&\quad + \begin{bmatrix} 0.512 & 0 \\ 0.224 & 0.064 \end{bmatrix} \begin{bmatrix} \epsilon_{1t-3} \\ \epsilon_{2t-3} \end{bmatrix} \\
&= \begin{bmatrix} 0.336\,2 \\ 1.735 \end{bmatrix} + \begin{bmatrix} 0.327\,7 & 0 \\ 0.158\,7 & 1.024 \times 10^{-2} \end{bmatrix} \begin{bmatrix} x_{t-5} \\ y_{t-5} \end{bmatrix} \\
&\quad + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} + \begin{bmatrix} .8 & 0 \\ .2 & .4 \end{bmatrix} \begin{bmatrix} \epsilon_{1t-1} \\ \epsilon_{2t-1} \end{bmatrix} + \begin{bmatrix} 0.64 & 0 \\ 0.24 & 0.16 \end{bmatrix} \begin{bmatrix} \epsilon_{1t-2} \\ \epsilon_{2t-2} \end{bmatrix} \\
&\quad + \begin{bmatrix} 0.512 & 0 \\ 0.224 & 0.064 \end{bmatrix} \begin{bmatrix} \epsilon_{1t-3} \\ \epsilon_{2t-3} \end{bmatrix} + \begin{bmatrix} 0.409\,6 & 0 \\ 0.192 & 0.025\,6 \end{bmatrix} \begin{bmatrix} \epsilon_{1t-4} \\ \epsilon_{2t-4} \end{bmatrix}
\end{aligned}
$$

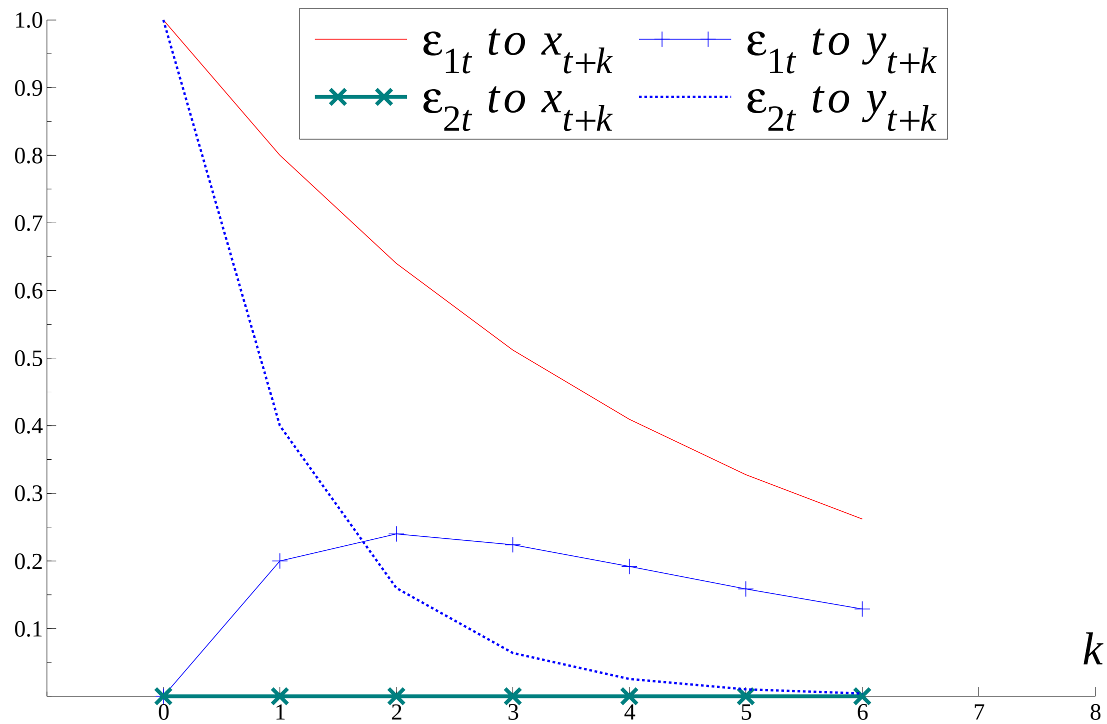
and finally

$$
\begin{aligned}
\begin{bmatrix} x_t \\ y_t \end{bmatrix} &= \begin{bmatrix} 0.369\,0 \\ 1.761 \end{bmatrix} + \begin{bmatrix} 0.262\,1 & 0 \\ 0.129 & 4.096 \times 10^{-3} \end{bmatrix} \begin{bmatrix} x_{t-6} \\ y_{t-6} \end{bmatrix} \\
&\quad + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} + \begin{bmatrix} .8 & 0 \\ .2 & .4 \end{bmatrix} \begin{bmatrix} \epsilon_{1t-1} \\ \epsilon_{2t-1} \end{bmatrix} + \begin{bmatrix} 0.64 & 0 \\ 0.24 & 0.16 \end{bmatrix} \begin{bmatrix} \epsilon_{1t-2} \\ \epsilon_{2t-2} \end{bmatrix} \\
&\quad + \begin{bmatrix} 0.512 & 0 \\ 0.224 & 0.064 \end{bmatrix} \begin{bmatrix} \epsilon_{1t-3} \\ \epsilon_{2t-3} \end{bmatrix} + \begin{bmatrix} 0.409\,6 & 0 \\ 0.192 & 0.025\,6 \end{bmatrix} \begin{bmatrix} \epsilon_{1t-4} \\ \epsilon_{2t-4} \end{bmatrix} \\
&\quad + \begin{bmatrix} 0.327\,7 & 0 \\ 0.158\,7 & 1.024 \times 10^{-2} \end{bmatrix} \begin{bmatrix} \epsilon_{1t-5} \\ \epsilon_{2t-5} \end{bmatrix} \\
&= \begin{bmatrix} 0.295\,2 \\ 1.69 \end{bmatrix} + \begin{bmatrix} 0.209\,7 & 0 \\ 0.104 & 1.638 \times 10^{-3} \end{bmatrix} \begin{bmatrix} x_{t-7} \\ y_{t-7} \end{bmatrix} \\
&\quad + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} + \begin{bmatrix} .8 & 0 \\ .2 & .4 \end{bmatrix} \begin{bmatrix} \epsilon_{1t-1} \\ \epsilon_{2t-1} \end{bmatrix} + \begin{bmatrix} 0.64 & 0 \\ 0.24 & 0.16 \end{bmatrix} \begin{bmatrix} \epsilon_{1t-2} \\ \epsilon_{2t-2} \end{bmatrix} \\
&\quad + \begin{bmatrix} 0.512 & 0 \\ 0.224 & 0.064 \end{bmatrix} \begin{bmatrix} \epsilon_{1t-3} \\ \epsilon_{2t-3} \end{bmatrix} + \begin{bmatrix} 0.4096 & 0 \\ 0.192 & 0.025\,6 \end{bmatrix} \begin{bmatrix} \epsilon_{1t-4} \\ \epsilon_{2t-4} \end{bmatrix} \\
&\quad + \begin{bmatrix} 0.3277 & 0 \\ 0.1587 & 1.024 \times 10^{-2} \end{bmatrix} \begin{bmatrix} \epsilon_{1t-5} \\ \epsilon_{2t-5} \end{bmatrix} + \begin{bmatrix} 0.2621 & 0 \\ 0.129 & 4.096 \times 10^{-3} \end{bmatrix} \begin{bmatrix} \epsilon_{1t-6} \\ \epsilon_{2t-6} \end{bmatrix}
\end{aligned}
$$

so the impact of the shock ϵ_{1t} to $(x_t, y_t) \dots (x_{t+6}, y_{t+6})$ is

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} .8 & 0 \\ .2 & .4 \end{bmatrix}, \begin{bmatrix} 0.64 & 0 \\ 0.24 & 0.16 \end{bmatrix}, \begin{bmatrix} 0.512 & 0 \\ 0.224 & 0.064 \end{bmatrix}, \begin{bmatrix} 0.4096 & 0 \\ 0.192 & 0.0256 \end{bmatrix}$$

$$\begin{bmatrix} 0.3277 & 0 \\ 0.1587 & 1.024 \times 10^{-2} \end{bmatrix}, \begin{bmatrix} 0.2621 & 0 \\ 0.129 & 4.096 \times 10^{-3} \end{bmatrix}$$



The graph shows that x_t is unaffected by shocks ϵ_{2t}, but that the impact of ϵ_{1t} is immediate and decays progressively. This is also the case for the response of y_t to ϵ_{2t}. By contrast, the response of y_t to an ϵ_{1t} shock builds up progressively and reaches a maximum after two periods (to a value of $0.24\epsilon_{2t}$) and then decays thereafter.

3. Does x_t Granger cause y_t, and vice versa?
ANSWER: We see from the VAR(1) representation that y_t does not Granger cause x_t but that x_t Granger causes y_t.

EXERCISE 12: VAR

Consider the following VAR model

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \begin{bmatrix} 0 & 1/2 \\ 1 & 1/2 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} u_t \\ v_t \end{bmatrix} \tag{6}$$

also denoted

$$\begin{aligned} \mathbf{z}_t &= \mu + \mathbf{A}\mathbf{z}_{t-1} + \epsilon_t \\ \mathbf{\Phi}(L)\,\mathbf{z}_t &= \mu + \epsilon_t \end{aligned}$$

with ϵ_t a white noise.

1. Check whether $\mathbf{z}_t$ is integrated of order 1
ANSWER:
$$\Phi(L) = \mathbf{I}_2 - \begin{bmatrix} 0 & 1/2 \\ 1 & 1/2 \end{bmatrix} L$$
and
$$\begin{aligned} |\Phi(z)| &= \begin{vmatrix} 1 & -1/2z \\ -z & 1-1/2z \end{vmatrix} = (1-1/2z) - z(1/2z) = -\frac{1}{2}\left(z^2+z-2\right) \\ &= \frac{1}{2}\left(z - \frac{-1+3}{2}\right)\left(z - \frac{-1-3}{2}\right) = \frac{1}{2}(z-1)(z+2) \end{aligned}$$
the two roots are $(z_1, z_2) = (1, -2)$ so the process is non stationary (given that there is only one unit root, we assume that it is integrated of order one, but this would need to be proven).

2. Compute $\mathbf{\Phi}(1)$: what is its rank? what do you conclude from this?
ANWER:
$$\mathbf{\Phi}(1) = \begin{bmatrix} 1 & -1/2 \\ -1 & 1/2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} [1 \qquad -1/2]$$
The rank of $\mathbf{\Phi}(1)$ is one (reduced rank). Notice that
$$\Delta \mathbf{z}_t = \mu - \mathbf{\Phi}(1)\,\mathbf{z}_{t-1} + \epsilon_t$$
so the process is integrated-cointegrated: there is one relation of cointegration

3. Write the Error Correction form for this model
$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} - \begin{bmatrix} 1 \\ -1 \end{bmatrix} (x_{t-1} - 1/2y_{t-1}) + \epsilon_t \tag{7}$$
i.e.
$$\begin{cases} \Delta x_t = -\left(x_{t-1} - 1/2y_{t-1} - 1\right) + u_t \\ \Delta y_t = \left(x_{t-1} - 1/2y_{t-1} - 1\right) + v_t \end{cases}$$

4. How many stochastic trends does $\mathbf{z}_t$ exhibit?
ANSWER: Define $\alpha = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $\beta = \begin{bmatrix} 1 \\ -1/2 \end{bmatrix}$, then we want to find matrices $\alpha_\perp$ and $\beta_\perp$ such that $\alpha_\perp' \alpha = \beta_\perp' \beta = 0$. We see that we can choose $\alpha_\perp = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\beta_\perp = \begin{bmatrix} 1/2 \\ 1 \end{bmatrix}$. Then multiplying (7) by β' yields
$$\begin{aligned} \Delta(\beta' \mathbf{z}_t) &= \Delta(x_{t-1} - 1/2y_{t-1}) = 3/2 - 3/2\,(x_{t-1} - 1/2y_{t-1}) + (u_t - 1/2v_t) \\ (x_t - 1/2y_t) &= 3/2 - 1/2\,(x_{t-1} - 1/2y_{t-1}) + (u_t - 1/2v_t) \end{aligned}$$
$x_t - 1/2y_t$ is stationary. Also multiplying (7) by $\alpha_\perp'$ yields
$$\begin{aligned} \Delta(\alpha_\perp' \mathbf{z}_t) &= \Delta(x_t + y_t) = u_t + v_t \\ (x_t + y_t) &= (x_{t-1} + y_{t-1}) + (u_t + v_t) \end{aligned}$$
$x_t + y_t$ follows a random walk (this is the stochastic trend that drives the non-stationary behavior in $\mathbf{z}_t$).

EXERCISE 13

What does it mean for two variables to be cointegrated? why is it problematic? why is it interesting?

EXERCISE 14: Volatility

Consider an ARCH(1) process given by

$$
\begin{aligned}
r_t &= \mu_t + \epsilon_t \\
\mu_t &= \phi_0 + \phi_1 r_{t-1} \\
\epsilon_t &\sim \mathsf{N}\left(0, \sigma_t^2\right) \\
\sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2
\end{aligned}
$$

1. What is $\mathsf{E}[r_t]$?
   ANSWER:

$$\mathsf{E}\left[r_t\right] = \mathsf{E}\left[\mu_t\right] = \phi_0 + \phi_1 \mathsf{E}\left[r_{t-1}\right]$$

   so if $|\phi_1| < 1$ (we implicitely assume an infinite history, so we can solve recursively upon stationarity)

$$\mathsf{E}\left[r_t\right] = \frac{\phi_0}{1-\phi_1}$$

2. What is $\mathsf{E}_{t-1}\left[r_t\right]$?
   ANSWER: Define $e_t = \epsilon_t / \sigma_t$ which is iid $\mathsf{N}(0,1)$.

$$
\begin{aligned}
\mathsf{E}_{t-1}\left[r_t\right] &= \mathsf{E}_{t-1}\left[\mu_t + \epsilon_t\right] \\
&= \mathsf{E}_{t-1}\left[\phi_0 + \phi_1 r_{t-1}\right] \\
&\quad + \mathsf{E}_{t-1}\left[\sigma_t e_t\right] \\
&= \phi_0 + \phi_1 r_{t-1} + \sqrt{\omega + \alpha_1 \epsilon_{t-1}^2}\, \mathsf{E}_{t-1}\left[e_t\right]
\end{aligned}
$$

   since e_t is iid $\mathsf{E}_{t-1}\left[e_t\right] = \mathsf{E}\left[e_t\right] = 0$, hence

$$\mathsf{E}_{t-1}\left[r_t\right] = \phi_0 + \phi_1 r_{t-1}$$

3. What is $\mathsf{E}\left[\epsilon_t^2\right]$?
   ANSWER:

$$
\begin{aligned}
\mathsf{E}\left[\epsilon_t^2\right] &= \mathsf{E}\left[\sigma_t^2 e_t^2\right] \\
&= \mathsf{E}\left[\mathsf{E}_{t-1}\left[\sigma_t^2 e_t^2\right]\right] \\
&= \mathsf{E}\left[\left(\omega + \alpha_1 \epsilon_{t-1}^2\right) \mathsf{E}_{t-1}\left[e_t^2\right]\right] \\
&= \mathsf{E}\left[\left(\omega + \alpha_1 \epsilon_{t-1}^2\right) \mathsf{E}\left[e_t^2\right]\right] \\
&= \mathsf{E}\left[\omega + \alpha_1 \epsilon_{t-1}^2\right] \\
&= \omega + \alpha_1 \mathsf{E}\left[\epsilon_{t-1}^2\right]
\end{aligned}
$$

   and again if $\alpha_1 \in (0,1)$

$$\mathsf{E}\left[\epsilon_t^2\right] = \frac{\omega}{1-\alpha_1}$$

4. What is $\mathsf{E}_{t-1}\left[\epsilon_t^2\right]$?
ANSWER:

$$\mathsf{E}_{t-1}\left[\epsilon_t^2\right]=\omega+\alpha_1\epsilon_{t-1}^2$$

5. What is $\mathsf{Cov}[\epsilon_t,\epsilon_{t-1}]$?
ANSWER:

$$\begin{aligned}
\mathsf{Cov}\left[\epsilon_t,\epsilon_{t-1}\right] &= \mathsf{E}\left[\epsilon_t\epsilon_{t-1}\right]\\
&= \mathsf{E}\left[\sigma_t e_t\sigma_{t-1}e_{t-1}\right]\\
&= \mathsf{E}\left[\mathsf{E}_{t-1}\left[\sigma_t e_t\sigma_{t-1}e_{t-1}\right]\right]\\
&= \mathsf{E}\left[\mathsf{E}_{t-1}\left[\sigma_{t-1}e_{t-1}\sqrt{\omega+\alpha_1\epsilon_{t-1}^2}e_t\right]\right]\\
&= \mathsf{E}\left[\left[\sigma_{t-1}e_{t-1}\sqrt{\omega+\alpha_1\epsilon_{t-1}^2}\right]\mathsf{E}_{t-1}\left[e_t\right]\right]\\
&= \mathsf{E}\left[\left[\sigma_{t-1}e_{t-1}\sqrt{\omega+\alpha_1\epsilon_{t-1}^2}\right]\mathsf{0}\right]=0
\end{aligned}$$

6. How does the conditional variance evolve in a GARCH(1,1)? Describe the evolution both mathematically and in words.
ANSWER: in a GARCH$(1,1)$, the conditional variance $\sigma_t^2=V\left[r_t|\mathcal{I}_{t-1}\right]$ follows the recursion

$$\sigma_t^2=\omega+\alpha\epsilon_{t-1}^2+\beta\sigma_{t-1}^2$$

where $\epsilon_t=r_t-E\left[r_t|\mathcal{I}_{t-1}\right]$. This shows that the conditional variance, i.e. the expected volatility, is a weighted average of both its value in the previous period (σ_{t-1}^2 is expected volatility at $t-2$) and the surprise in volatility, or unanticipated variance ($\epsilon_{t-1}^2-\sigma_{t-1}^2$, since $E_{t-2}\left(\epsilon_{t-1}^2-\sigma_{t-1}^2\right)=0$):

$$\sigma_t^2=\omega+\left(\alpha+\beta\right)\sigma_{t-1}^2+\alpha\left(\epsilon_{t-1}^2-\sigma_{t-1}^2\right)$$

7. What is the h-step ahead forecast formula for an ARCH(1) model?
ANSWER: From the ARCH(1) , $\sigma_t^2=\omega+\alpha\epsilon_{t-1}^2$, define $\nu_t=\epsilon_t^2-\sigma_t^2$, then $\mathrm{E}_{t-1}\left(\nu_t\right)=0$ and ν_t is called a *martingale difference sequence*, i.e. a process whose expectation conditional on the previous period's information set is zero. It follows that

$$\epsilon_t^2=\omega+\alpha\epsilon_{t-1}^2+\nu_t$$

so ϵ_t^2 follows an AR(1) . Its h-step forecast is given by the conditional expectation

$$\begin{aligned}
E_t\epsilon_{t+h}^2 &= \omega+E_t\alpha\epsilon_{t+h-1}^2\\
&= ...=\omega\sum_{i=0}^{h-1}\alpha^i+\alpha^h\epsilon_t^2\\
&= \omega\frac{1-\alpha^h}{1-\alpha}+\alpha^h\epsilon_t^2
\end{aligned}$$

and since, using the law of iterated expectations

$$E_t\left(\epsilon_{t+h}^2\right)=E_t\left(E_{t+h-1}\left(\epsilon_{t+h}^2\right)\right)=E_t\left(\sigma_{t+h}^2\right)$$

it follows that the forecast $\sigma^2_{t+h|t} \overset{def}{=} E_t \sigma^2_{t+h}$

$$\sigma^2_{t+h|t} = \omega \frac{1-\alpha^h}{1-\alpha} + \alpha^h \epsilon^2_t$$

8. What is the h-step ahead forecast formula for a GARCH(1,1) model?
ANSWER: for the GARCH$(1,1)$

$$\epsilon^2_t = \omega + (\alpha + \beta)\, \epsilon^2_{t-1} + \nu_t - \beta \nu_{t-1}$$

so from the formula for an ARMA$(1,1)$, the forecast differs from that of an AR(1) only at horizon $h = 1$:

$$\begin{aligned} h &= 1 : E_t \sigma^2_{t+1} = \omega + (\alpha+\beta)\, \epsilon^2_t - \beta \nu_t \\ h &\geq 2 : E_t \sigma^2_{t+h} = \omega \frac{1-(\alpha+\beta)^h}{1-(\alpha+\beta)} + (\alpha+\beta)^h\, \epsilon^2_t \end{aligned}$$

9. What is the main advantage of a GARCH process over an ARCH process?
ANSWER: because the GARCH allows for an ARMA representation for ϵ^2_t whereas the ARCH only provides an AR, the GARCH is more flexible and requires fewer parameters to represent persistent dynamics where long lags matter (since an MA(1) can be inverted into an AR(∞)).

10. Name and describe one or two important features the classic GARCH model is missing?
ANSWER:

    - the classic GARCH lacks asymmetry. This is an empirical feature of many financial variables, whereby positive shocks have a lower impact on volatility than negative shocks.
    - Also, the GARCH model imposes that the conditional variance follows an ARMA, whereas, it might be the conditional standard deviation, or another power.
    - Volatility in GARCH models is driven by the suprise in returns, i.e. ϵ_t. It might be the case that volatility is an entirely autonomous process that is correlated with ϵ_t but not entirely driven by it. To relax the latter assumption is the purpose of Stochastic Volatility models.