Lecture 2: AR models

Dr. Joao B. Duarte[1]

[1]Nova School of Business and Economics
University of Cambridge

Masters, Economics: Macroeconometrics
Lisbon
Spring 2017

Lecture Objectives:

- Introduce stochastic difference equations and the concept of stationarity.
- Describe the AR models and their properties.
- Show the stationarity conditions for AR models and how they are linked with the stability conditions of a SDE.
- Define the autocorrelation function (ACF) and the partial autocorrelation function (PACF).
- Show how they can help us identify an AR model, select its lag and check the residuals of an estimated AR model.
- Introduce the information criteria AIC and SBC.
- Introduce the Box-Jenkins methodology.

Secondary Readings:

- Chapter 2, Applied Econometric Time Series, Enders, Walter, Fourth Edition
- Chapter 3, Time Series Analysis, Hamilton, James, first edition

Stochastic Difference Equations

- Suppose we have again our benchmark examples:

$$y_t = a_1 y_{t-1} + \varepsilon_t \tag{1}$$

- If we now assume that $\varepsilon_t \sim N(0, \sigma^2)$
- Then (1) becomes a stochastic difference equation.
- In this course we will generally model the stochastic component as white noise.

White Noise - Definition

- A white noise process is a sequence $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ such that:

  1. $E(\varepsilon_t) = 0$
  2. $E(\varepsilon_t^2) = \sigma^2$
  3. $E(\varepsilon_t \varepsilon_h) = 0 \quad$ for $t \neq h$

- Moreover, if we add the assumption that ε_t is normally distributed:

  4. $\varepsilon_t \sim N(0, \sigma^2)$

- Then we have a Gaussian white noise.

Stochastic Processes

- Suppose we have observed a sample of size T of the random variable (1) with $a_1 = 0.5$.

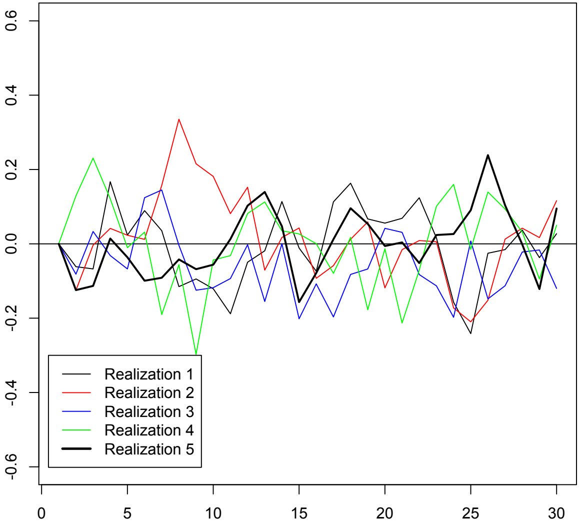
$$\{y_1, y_2, ..., y_T\} \tag{2}$$

- This sample is clearly dependent on the specific draws of the T independent and identically distributed ε_t:

$$\{\varepsilon_1, \varepsilon_2, ..., \varepsilon_T\} \tag{3}$$

- Hence, the observed sequence is only one realization of the stochastic difference equation.

Many Realizations of the same Stochastic Process



Stochastic Process

- Now suppose we take a battery of I sequences such that at each date t we would have I observations coming from the same distribution:

$$\{y_t^{(1)}, y_t^{(2)}, ..., y_t^{(I)}\}$$

- This random variable, call it Y_t, for a Gaussian white process has the following density:

$$f_{Y_t}(y_t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-y_t^2}{2\sigma^2}\right]$$

Stochastic Process - Unconditional Mean

- If the unconditional mean exists, at time t, it would be given by:

$$E(Y_t) = \int_{-\infty}^{\infty} y_t f_{Y_t}(y_t) dy_t \tag{4}$$

- Example: For a Gaussian white noise process:

$$Y_t = \mu + \varepsilon_t \tag{5}$$

$$E(Y_t) = \mu \tag{6}$$

Why?

Stochastic Process - Autocovariance

- The realizations can have time dependence
- We can use the joint distribution to understand this dependence through the autocovariance.
- The autocovariance can be calculated from the joint distribution of $(Y_t, Y_{t-1}, ...)$

$$\gamma_{jt} = E(Y_t - \mu_t)E(Y_{t-j} - \mu_{t-j}) \tag{7}$$

- For Example (5) the autocovariance is:

$$\begin{aligned} \gamma_{0t} &= E(Y_t - \mu_t)^2 = \sigma^2 \\ \gamma_{jt} &= 0 \quad j \neq 0 \end{aligned}$$

Stationarity

- Time series objective is to model the time dependence of observations.
- Observations can only be independent in one way but they can be dependent in many ways.
- One important way to model time dependence is with stationary models (Later in the course we will discuss non-stationary models).
- The foundation of statistical inference in time series and forecast is the concept of weak stationarity.
- ARMA are very useful stationary models for univariate time series.

Stochastic Process - Weakly Stationary

- Definition: A stochastic process Y_t is weakly stationary if neither the mean μ_t nor the autocovariances γ_{jt} depend on date t:

$$\begin{aligned} E(Y_t) &= \mu \quad \forall t && (8) \\ E(Y_t - \mu)E(Y_{t-j} - \mu) &= \gamma_j \quad \forall t, j && (9) \end{aligned}$$

- A Gaussian process that is weakly stationary is also strictly stationary.
- Verify (5) is weakly stationary.

Examples - Is it stationary?

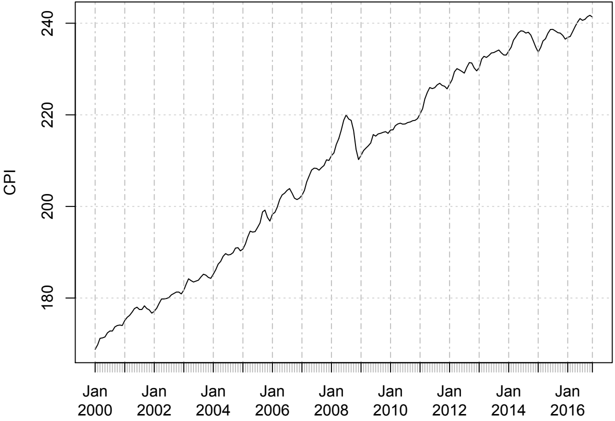


Figure: U.S. Consumer Price Index

Examples - Is it stationary?

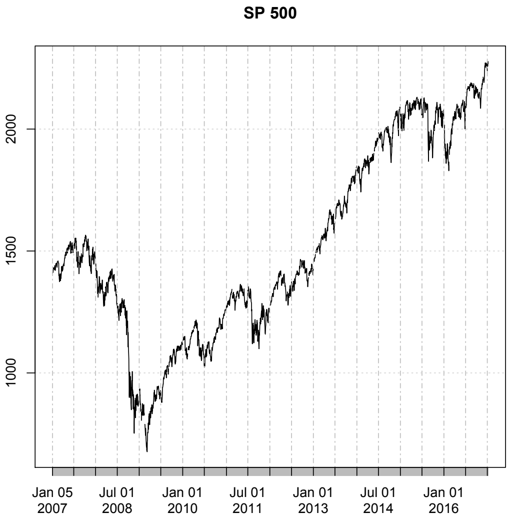


Figure: SP 500 U.S. Stock Index

Examples - Is it stationary?

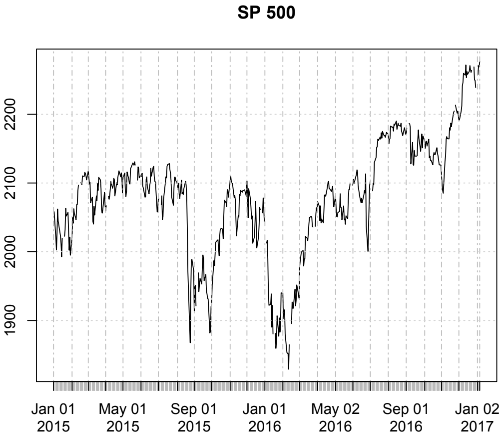


Figure: SP 500 U.S. Stock Index

Examples - Is it stationary?

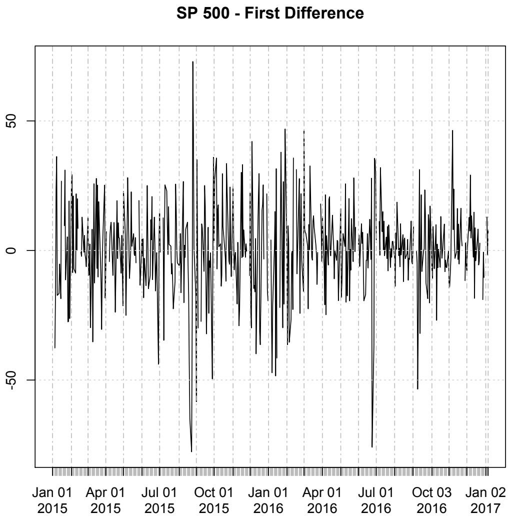


Figure: SP 500 U.S. Stock Index - First Difference

Examples - Is it stationary?

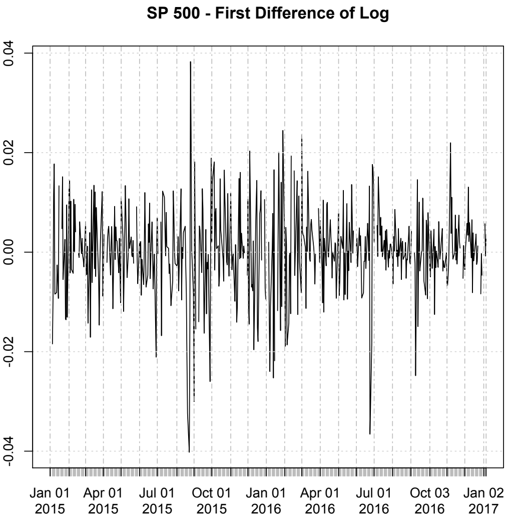


Figure: SP 500 U.S. Stock Index - First Difference of the Log

Examples - Is it stationary?

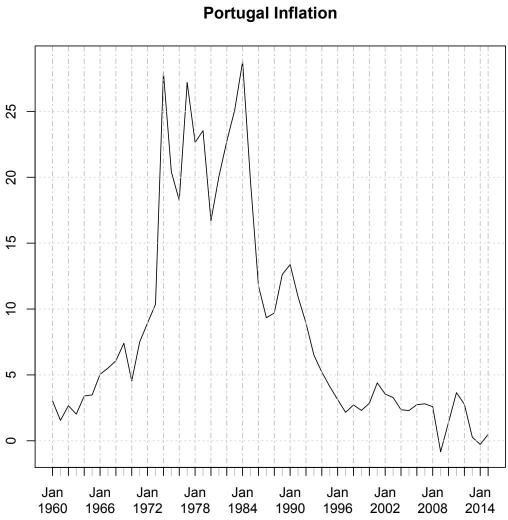


Figure: Portugal Yearly Inflation

Stochastic Process - Weakly Stationary

- Unfortunately, we do not observe I realizations. We just get one. Hence, we cannot be sure that the time average:

$$\bar{y} = (1/T)\sum_{t=1}^{T} y_t^{(1)} = (1/I)\sum_{i=1}^{I} y_t^{(i)} \tag{10}$$

- *It turns out that under very general conditions, a weakly stationary process is ergodic for the mean.*
- Which means (10) holds and we can use standard econometric techniques using our sample of observations.

Autoregressive Models - AR(1)

- An AR model of order 1, or simply AR(1) uses the first lagged observation to predict the current observation:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t \tag{11}$$

where ε_t is assumed to be a white noise process.

- The conditional expectation of y_t is:

$$E(y_t|y_{t-1}) = \phi_0 + \phi_1 y_{t-1} \tag{12}$$

AR Properties

- Is the AR(1) process stationary?
- What if $|\phi_1| > 1$? Example: $\phi_1 = 1.1$

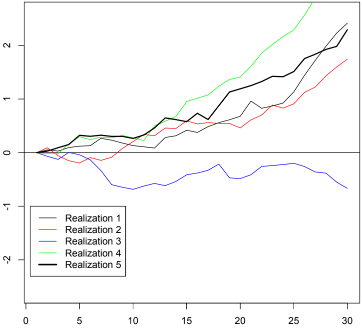


Figure: Realizations with $\phi_1 = 1.1$

AR Properties

- Hence, there is a close connection between the stability of a difference equation and stationarity of a stochastic process.
- Clearly, a stochastic difference equation cannot be stationary if $|\phi_1| > 1$.
- Moreover, it turns out that an AR(1) process is weakly stationary if and only if $|\phi_1| < 1$.
- We will prove here just one direction, from weakly stationary to $|\phi_1| < 1$. Please see notes and/or Hamilton for the other direction of the proof.

AR Properties

- Assuming the series is weakly stationary: $E(y_t) = \mu$, $Var(y_t) = \gamma_0$ and $Cov(y_t, y_{t-j}) = \gamma_j$.

  Taking expectation on (11) we have:

$$E(y_t) = \phi_0 + \phi_1 E(y_{t-1}) + E(\varepsilon_t) \Rightarrow$$

$$E(y_t) = \phi_0 + \phi_1 E(y_{t-1}) \Rightarrow$$

$$\mu = \phi_0 + \phi_1 \mu \Rightarrow$$

$$\boxed{\mu = \frac{\phi_0}{1 - \phi_1}} \tag{13}$$

AR Properties

- Two interesting notes:
  1. The unconditional mean of y_t only exists if $\phi_1 \neq 1$
  2. The mean of y_t, $\mu = 0$ iff (if and only if) $\phi_0 = 0$

- Rearranging (13) $\Rightarrow \phi_0 = (1 - \phi_1)\mu$, we can substitute it in (11) to get:

$$y_t - \mu = \phi_1(y_{t-1} - \mu) + \varepsilon_t \tag{14}$$

- If we take the square and then the expectation of (14) we obtain the variance:

$$E(y_t - \mu)^2 = Var(y_t) = \phi_1^2 E(y_{t-1} - \mu)^2 + 2E[\phi_1(y_{t-1} - \mu)\varepsilon_t] + E(\varepsilon_t)^2$$

AR Properties

- Since $E(\varepsilon_t) = 0$ (why?) and $Cov(y_{t-1}, \varepsilon_t) = E[(y_{t-1} - \mu)\varepsilon_t] = 0$ (why?)

$$Var(y_t) = \phi_1^2 Var(y_{t-1}) + \sigma^2$$

- Because we assumed stationarity, we have that $Var(y_t) = Var_{t-1}$

$$\boxed{Var(y_t) = \frac{\sigma^2}{1 - \phi_1^2}} \tag{15}$$

AR Properties

- Since variance of a random variable is nonnegative we have that $\phi_1^2 < 1 \Rightarrow |\phi_1| < 1$.
- Now we move to show the autocovariance of the AR(1) process. Multiplying (14) by $(y_{t-j} - \mu)$:

$$\gamma_j = \phi_1 \gamma_{j-1} \quad if \quad j > 0 \tag{16}$$

- and $\gamma_0 = Var(y_t)$

Autocorrelation

- Definition: The autocorrelation is given by:

$$\rho_j = \frac{Cov(y_t, y_{t-j})}{\sqrt{Var(y_t)Var(y_{t-j})}} \tag{17}$$

- Given stationarity, $Cov(y_t, y_{t-j}) = \gamma_j$ and $Var(y_t) = Var(y_{t-j}) \Rightarrow$

$$\rho_0 = 1 \tag{18}$$

$$\rho_j = \phi_1 \rho_{j-1} \quad if \quad j > 0 \tag{19}$$

- The autocorrelation function (ACF) of an AR(1) is given by $\rho_j = \phi_1^j$

AR(2)

- The AR(2) process assumes the form

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \tag{20}$$

- Using the same technique as we used in AR(1):

$$E(y_t) = \boxed{\mu = \frac{\phi_0}{1 - \phi_1 - \phi_2}} \tag{21}$$

AR(2) - Properties

- Provided $\phi_1 + \phi_2 \neq 1$. We can use $\phi_0 = (1 - \phi_1 - \phi_2)\mu$ we obtain:

$$(y_t - \mu) = \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + \varepsilon_t \qquad (22)$$

- Multiplying (22) by $(y_{t-j} - \mu)$ we have:

$$\begin{aligned}(y_{t-j} - \mu)(y_t - \mu) &= \phi_1(y_{t-j} - \mu)(y_{t-1} - \mu) \\ &\quad + \phi_2(y_{t-j} - \mu)(y_{t-1} - \mu) + (y_{t-j} - \mu)\varepsilon_t\end{aligned}$$

- Taking Expectation

$$\gamma_j = \phi_1\gamma_{j-1} + \phi_2\gamma_{j-2} \quad \forall j > 0 \qquad (23)$$

AR(2) - Properties

- Exercise: Find the variance of $y_t = \gamma_0$.
- (23) is a homogeneous difference equation. You can see the solution in the last class lecture.
- Here we use the Lag operator to show an alternative way of solving (23).

$$\gamma_j = \phi_1 L \gamma_j + \phi_2 L^2 \gamma_j$$

$$(1 - \phi_1 L - \phi_2 L^2)\gamma_j = 0 \tag{24}$$

- There is a correspondence between (24) and the following polynomial (see Hamilton for details):

$$1 - \phi_1 z - \phi_2 z^2 = 0 \tag{25}$$

AR(2) - Properties

- There is also a direct correspondence with the characteristic roots we saw in Lecture 1 and the roots of the (25) polynomial:

$$z_1 = \alpha_1^{-1} \tag{26}$$
$$z_2 = \alpha_2^{-1} \tag{27}$$

- Point: We can alternatively find the homogeneous solution by calculating the roots of (25).

$$(z_1, z_2) = \left(\frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2})}{-2\phi_2}, \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2})}{-2\phi_2} \right)$$

AR(2) - Stationarity

- The stationarity condition is again closely connected with the stability condition of the difference equation.
- The stationarity condition for an AR(2) process is that the roots z_1, z_2 lie outside the unit circle.
- Alternatively, the stationarity condition for an AR(2) process is that the roots α_1, α_2 lie inside the unit circle.

AR(p)

- The results of AR(1) and AR(2) can readily be extended to the general AR(p) model

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \tag{28}$$

$$\boxed{E(y_t) = \frac{\phi_0}{1 - \phi_1 - \dots - \phi_p}} \tag{29}$$

- The associated characteristic equation is

$$1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p = 0 \tag{30}$$

- If all roots are greater than 1 in modulus, y_t is stationary. Again, inverses of the roots are the characteristic roots of the model.
- Hence, y_t is stationary if the characteristic roots of the model are less than 1 in modulus.

Identifying AR(p) in practice - Box-Jenkins Methodology

- **Stationarity**
- **Order Determination**
  1. Partial Autocorrelation Function (PACF)
  2. Information Criteria
- **Parameter Estimation**
- **Model Checking**
- **Forecast**

PACF

- Consider the following AR models in consecutive order:

$$
\begin{aligned}
y_t &= \phi_{0,1} + \phi_{1,1} y_{t-1} + \varepsilon_{1t} \\
y_t &= \phi_{0,2} + \phi_{1,2} y_{t-1} + \phi_{2,2} y_{t-2} + \varepsilon_{2t} \\
y_t &= \phi_{0,3} + \phi_{1,3} y_{t-1} + \phi_{2,3} y_{t-2} + \phi_{3,3} y_{t-3} + \varepsilon_{3t} \\
&\dots
\end{aligned}
$$

- Since, ε are white noise, we can use least squares regression to estimate these models.

- The lag-1 sample PACF of y_t is the estimate $\hat{\phi}_{1,1}$. The lag-2 sample PACF is given by $\hat{\phi}_{2,2}$ and so on.

- Hence, the lag-l PACF measures the added contribution of lag l to the AR($l-1$) model.

PACF

- Hence, we select the lag based on when the PACF turns to zero. In other words, when we choose the lag based on the lag upon which there is no more added contribution.
- Example: Portugal GDP first difference. We choose an AR(4)

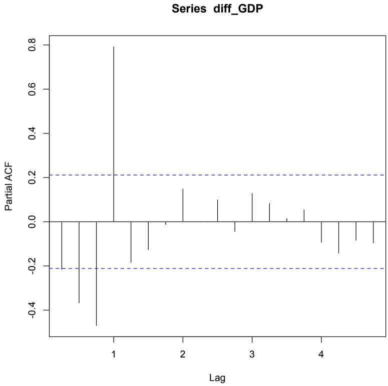


Figure: Portugal's GDP first difference PACF

Information Criteria

- There are several information criteria available to determine the order p of an AR process.
- For a Gaussian process, the Akaike Information Criterion (AIC) is the following:

$$AIC = \ln\left(\hat{\sigma_p^2}\right) + \frac{2p}{T} \tag{31}$$

- Another commonly used criteria is the Schwarz-Bayesian criteria (BIC, Bayesian Information Criteria):

$$BIC(p) = \ln\left(\hat{\sigma_p^2}\right) + \frac{p\ln(T)}{T} \tag{32}$$

- We compute the AIC and BIC for each lag and select the model lag that had the minimum AIC or BIC.

AIC and BIC

- Back to our Example: Portugal GDP first difference.

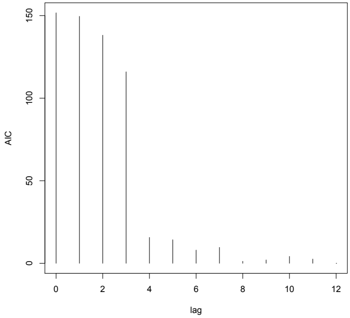


Figure: Portugal's GDP first difference AIC

- We would choose lag 12. However, the good practice in lag selection prefers smaller models. Since the AIC falls dramatically at lag 4, the best thing here would be to also choose lag 4 like the PACF selected.

Parameter Estimation

- For a specified AR(p) model, we can use the conditional least squares (LS) method which makes use of the (p+1)th observation.

- The fitted model is:

$$\hat{y}_t = \hat{\phi}_0 + \hat{\phi}_1 y_{t-1} + ... + \hat{\phi}_p y_{t-p} \tag{33}$$

  and the associated residual is

$$\hat{\varepsilon}_t = y_t - \hat{y}_t \tag{34}$$

- From the residuals we can estimate the variance of ε_t by:

$$\hat{\sigma}^2 = \frac{\sum_{t=p+1}^{T} \hat{\varepsilon}_t^2}{T - 2p - 1} \tag{35}$$

Parameter Estimation

- Example: Portugal first difference GDP estimation of AR(4):

Parameter	Estimate
$\hat{\phi}_0$	128.4608
	(101.3962)
$\hat{\phi}_1$	-0.0888
	(0.0560)
$\hat{\phi}_2$	-0.0983
	(0.0533)
$\hat{\phi}_3$	-0.0765
	(0.0565)
$\hat{\phi}_4$	0.8348
	(0.0537)
$\hat{\sigma}^2$	186434

Table: AR(4) estimates.

Model Checking

- If the model is adequate, the residual series should behave as a white noise.
- We can use the ACF of the residuals to check this and/or the Ljung-Box statistics.
- For an AR(p) model, the Ljung-Box statistic $Q(m)$ follows asymptotically a chi-squared distribution with $m - g$ degrees of freedom. g denotes the number of AR coefficients.

Model Checking

- Example: The ACF of the residuals is:

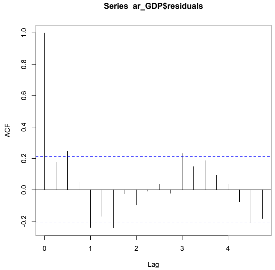


Figure: Portugal's GDP first difference ACF of residuals

- And the Ljung-Box statistic, $Q(12)$ is 28.67 and we reject the null hypothesis of no serial correlation in the residuals.

Model Checking

- Hence, the model is not adequate. Both the ACF and the Ljung-Box statistic give the same conclusion.
- The problem is the seasonality that is still present in the time series of Portugal GDP first differences.
- We will work on removing this seasonality in the first work assignment.

Summary

- Stationary processes allow to use the standard statistics toolkit.
- The mainly reason being that a stationary process is generally ergodic for the mean.
- An AR process is stationary if all of its characteristics roots lie inside the unit circle.
- The PACF can be useful for determining the number of lags in an AR model. So are the AIC and BIC information criteria methods.
- After estimating the AR model, we should check the residuals to see that they are stationary.

Questions to think about

- Are AR models flexible enough? Can they capture most time series behaviours?
- Why does the ACF decay slowly in an AR model?
- How should we interpret the lag parameters in an AR model?