Lecture 11: ARIMA and SARIMA models

Josep Vives

April 20, 2024

1 ARIMA models

Let $\{Y_j,\ j \geq 0\}$ be a random walk. We saw this process at the beginning of the course. It is defined as

$$Y_j = Y_{j-1} + Z_j,$$

where $Z \sim WN(0, \sigma^2)$ and $Y_0 = Z_0 = 0$. We know that this is not a stationary process because

$$\mathbb{V}(Y_j) = \mathbb{V}(Y_{j-1}) + \sigma^2$$

and so,

$$\mathbb{V}(Y_j) = j\,\sigma^2.$$

But note that if we compute its first differences we obtain

$$X_j := Y_j - Y_{j-1} = Z_j, \quad \forall j \geq 1$$

and so, the process of first differences is stationary.

This example illustrates that differencing can convert non-stationary series in stationary ones. This is the basis to generalize ARMA models to the family of ARIMA models that includes non-stationary models

Definition 1.1 *We say that X is an ARIMA(p,d,q) model with $p, d, q \in \mathbb{N}$ if*

$$Y_j = (\mathrm{Id} - B)^d\, X_j, \quad j \in \mathbb{Z}.$$

is an ARMA(p,q) model.

Process Y is process X differentiated d times and process X is process Y integrated d times. Integrated processes are called ARIMA processes. Letter I means *integrated*. Therefore, a process X is an ARIMA process if it satisfies

$$\Phi_p(B)\,(\mathrm{Id} - B)^d\, X_j = \Theta_q\,(B)\, Z_j, \quad j \in \mathbb{Z}.$$

Obviously, ARIMA(p,0,q) means ARMA(p,q). The random walk is an ARIMA(0,1,0) process. Of course, an ARIMA process can have a trend. In this case we have

$$\Phi_p(B)\,(\mathrm{Id} - B)^d\, X_j = \delta + \Theta_q\,(B)\, Z_j, \quad j \in \mathbb{Z}.$$

Indeed, note that for the particular case of $p = q = 0$ and $d = 1$ we have,

$$X_j = X_{j-1} + \delta + Z_j.$$

And therefore,

$$E(X_j) = E(X_{j-1}) + \delta = \cdots = j\,\delta.$$

2 Prices

An interesting example of ARIMA process is the following. Let $S = \{S_j\ j \geq 0\}$ be a price process in an open market, that is, under the offer and demand law. As a first attempt to model prices it is reasonable to assume

$$\text{Law}\left(\frac{S_j - S_{j-1}}{S_{j-1}}\right) = \text{Law}\left(\frac{S_{j+1} - S_j}{S_j}\right),$$

that is, relative increments are stationary, or in other words, the same law determine the relative price variation of today and the relative price variation of tomorrow. Then,

$$\text{Law}\left(\frac{S_j}{S_{j-1}}\right) = \text{Law}\left(\frac{S_{j+1}}{S_j}\right),$$

and therefore,

$$\text{Law}(\log S_j - \log S_{j-1}) = \text{Law}(\log S_{j+1} - \log S_j),$$

that is, $X_j := \log S_j - \log S_{j-1}$ is a stationary process.

Therefore, if X is an ARMA(p,q) process, then $\log S$ is an ARIMA(p,1,q) process, and it is also a logarithmic transformation of the time series of prices.

Note that if $|x| < 1$ we have the Taylor expansion

$$\log(1+x) = x + \frac{1}{2}\,x^2 + \frac{1}{3}\,x^3 - \cdots$$

and so, writing $y = 1 + x$, if $|y - 1| < 1$, we have

$$\log y \simeq y - 1,$$

$$\log \frac{S_j}{S_{j-1}} \simeq \frac{S_j}{S_{j-1}} - 1 = \frac{S_j - S_{j-1}}{S_{j-1}}$$

and

$$|\frac{S_j}{S_{j-1}} - 1| < 1 \Longleftrightarrow \left|\frac{S_j - S_{j-1}}{S_{j-1}}\right| < 1.$$

Recall that if $|\frac{S_j - S_{j-1}}{S_{j-1}}| > 0.2$, typically, markets close the trade of the asset.

Therefore, except on critical situations, increments of logarithmic prices coincide with relative increments of prices, that is, with returns. This is the reason why the process of logarithmic prices is of main importance in Finance.

3 Dickey and Fuller test of unitary roots

A typical problem in Time Series Analysis is to distinguish between a random walk and an AR(1) process with a parameter ϕ close to 1. Recall that the two processes are very different because one is stationary and the other not. More formally, the problem is the following: given

$$X_j = \phi X_{j-1} + Z_j, \ |\phi| \leq 1,$$

how to distinguish $\phi = 1$ from $\phi < 1$?

We can rewrite the model as

$$X_j - X_{j-1} = (\phi - 1) X_{j-1} + Z_j,$$

that is, if $\rho := \phi - 1$,

$$(Id - B) X_j = \rho X_{j-1} + Z$$

and we have to distinguish between the null hypothesis $\rho = 0$ and the alternative hypothesis $\rho < 0$.

This can be done by the statistic of Dickey-Fuller that under the null hypothesis has the expression

$$\hat{\rho} = DF_n := n \frac{\sum_{i=1}^{n} x_{i-1}(x_i - x_{i-1})}{\sum_{i=1}^{n} x_{i-1}^2}$$

and can be deduced immediately from the least squares estimator of ϕ.

Note that DF_n is less than 0, and near to 0 if the null hypothesis is true and far from it if the true hypothesis is the alternative. With a 5% of error, if DF_n is less than -1.94 we accept the alternative hypothesis that it is an AR(1) and if it is greater than -1.94, we accept null hypothesis that it is a random walk.

An alternative to Dickey-Fuller test, that runs similarly, is the Phillips-Perron test. The Dickey-Fuller test can be generalized allowing the error Z have a mean or a deterministic trend. Or being a stationary ARMA process. In this last case, the generalization is the so called Augmented Dickey-Fuller test, and it runs similarly.

4 SARIMA models

Let $\{X_j, \ j \geq 1\}$ be a time series. Recall that we can write it as $X_j = X_{i+(r-1)s}$ with $i = 1, \cdots, s$ and $r \geq 1$, where s is the period. For example, given a series of monthly data, $s = 12$ and r indicates the year. Therefore, the first figure is $i = 1$ and $r = 1$, that is, the month 1 of the year 1. The thirteen figure is $i = 1$ and $r = 2$, that is, $j = 1 + (2-1) \times 12 = 13$, and so on.

4.1 Pure seasonal models

To fix ideas assume a yearly seasonal period. A pure seasonal model assumes that there is correlation only between data of the same month, and data of different month are uncorrelated. So, in fact, we have 12 time series, one per month, uncorrelated between them.

If the model is ARMA(a, c) this means

$$X_{i+(r-1)s} - \alpha_1 X_{i+(r-2)s} - \cdots - \alpha_a X_{i+(r-a-1)s} = U_{i+(r-1)s} - \beta_1 U_{i+(r-2)s} - \cdots - \beta_c U_{i+(r-c-1)s}$$

with $\{U_{i+(r-1)s},\ r \geq 1, i = 1, \ldots, s\} \sim WN(0, \sigma^2)$. This model is denoted by ARMA $(a, c)_s \times WN(0, \sigma^2)$.

For example, a $\text{AR}(1)_{12} \times WN(0, \sigma^2)$ satisfies

$$X_{i+(r-1)12} - \alpha\, X_{i+(r-2)12} = U_{i+(r-1)12}, \quad |\alpha| < 1,$$

or equivalently,

$$X_j - \alpha\, X_{j-12} = U_j, \quad j \in \mathbb{Z}, \quad |\alpha| < 1.$$

Analogously to the non seasonal AR(1) case we have

$$\gamma(0) = \frac{\sigma^2}{1 - \alpha^2}$$

and

$$\rho(j) = \begin{cases} 1, & j = 0 \\ \alpha^l, & j = l \cdot s \\ 0, & \text{otherwise} \end{cases}$$

Note that we can write

$$(\text{Id} - \alpha\, B^{12})\, X_j = U_j, \quad j \in \mathbb{Z}$$

or

$$X_j = (\text{Id} - \alpha\, B^{12})^{-1}\, U_j, \quad j \in \mathbb{Z}.$$

4.2 Seasonal models

Think now on the model MA $(1) \times$ MA $(1)_{12}$. Its equation is

$$X_j = (\text{Id} - \theta\, B)(\text{Id} - \vartheta\, B^{12}) Z_j, \quad j \in \mathbb{Z}$$

with

$$Z \sim WN(0, \sigma^2),$$

that is, in a displayed form we have

$$X_j = Z_j - \vartheta\, Z_{j-12} - \theta\, Z_{j-1} + \theta\, \vartheta Z_{j-13}.$$

An $\text{AR}(1) \times \text{AR}(1)_{12}$ model would be

$$X_j = (\text{Id} - \phi\, B)^{-1} (\text{Id} - \varphi\, B^{12})^{-1} Z_j, \quad j \in \mathbb{Z}$$

and in a displayed form,

$$X_j - \varphi X_{j-12} - \phi X_{j-1} + \phi \varphi X_{j-13} = Z_j.$$

More general models are

$$\text{ARMA } (p,q) \times \text{ARMA } (a,c)_s,$$

or still more generally,

$$\text{SARIMA } (p,d,q)\,(a,D,c)_s.$$

The general equation of a SARIMA model is

$$\Phi_p(B)\,A_a(B^s)(\mathrm{Id}-B)^d(\mathrm{Id}-B^s)^D\,X_j = \mu + \Theta_q(B)\,C_c\,(B^s)\,Z_j$$

Applying this formula we can write some examples:

a) A $\text{SARIMA}(0,0,0)(1,0,0)_{12}$ is given by

$$X_j = \varphi X_{j-12} + W_j.$$

b) A $\text{SARIMA}(0,0,0)(0,1,1)_4$ is given by

$$X_j = X_{j-4} + W_j - \vartheta W_{j-4}.$$

c) A $\text{SARIMA}(0,1,0)(1,0,0)_{12}$ is given by

$$X_j = X_{j-1} + \varphi X_{j-12} - \varphi X_{j-13} + W_j.$$

d) A $\text{SARIMA}(0,1,0)(0,0,1)_4$ is given by

$$X_j = X_{j-1} + W_j - \vartheta W_{j-4}.$$