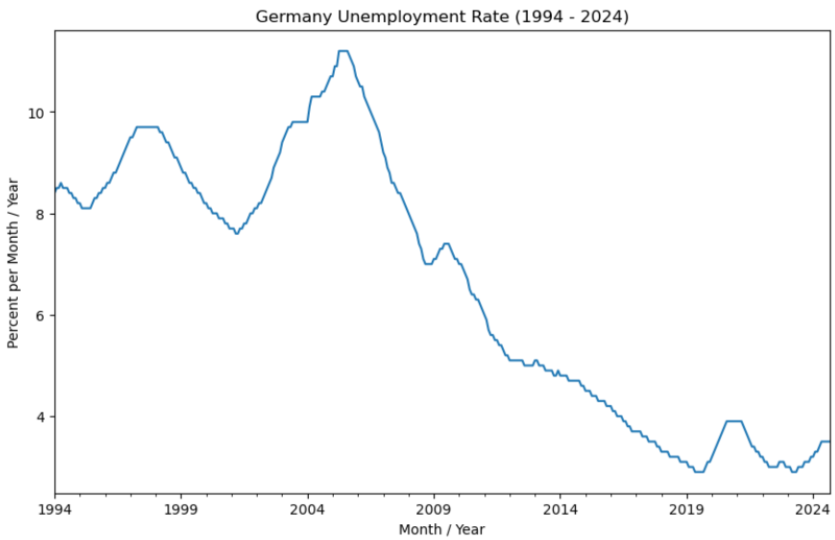
Time Series: A Summary

Artem Urlapov Sedova (Universidad Autónoma de Madrid)

A time series is an orderly representation of given values in time.

Example:



The common approach is to obtain insights by means of a statistical regression, of the type:

$y_t = a + b\, y_{t-1} + \varepsilon_t$, where:

y_t is the dependent variable

a is the intercept

b is a coefficient

y_{t-1} is the lagged explanatory variable

ε_t is the residual

If ε_t has mean = 0 and variance = σ^2 (or variance = 1 in a standard normal distribution), we call the residual *white noise*.

The motivation behind a univariate (one variable) or a multivariate (two or more variables) TS analysis, conditioned on data availability, has to do with the possibility of gaining more insights through multivariate time series analysis, as two or more variables can explain more than one variable.

However, there is a trade-off between model complexity and explanatory power: for this reason, a simpler, univariate time series analysis is sometimes preferred.

1. Univariate Time Series

ARIMA (p, d, q) is the usual approach towards modelling univariate TS. Thus, we have:

- AR component (p), which denotes the AutoRegressive part, this is to say, the regression of the variable on past values of the variable.
- Differencing order (d), which indicates the number of differences we need to take for the series to be stationary (in economics, d = 0, 1, 2; in other words, no more than 2).
- MA component (q), which denotes the Moving Averages part: the regression of the variable on past values of the residual.

Example: ARIMA (2, 0, 1). [Can also be represented as ARMA (2, 1).] Here, p = 2, d = 0, and q = 1. Mathematically, this is expressed as:

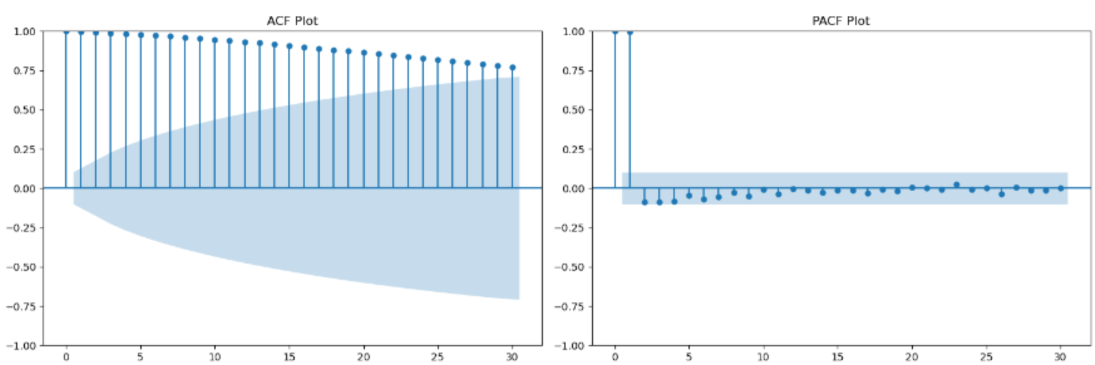
$$y_t = c + \phi_{1_{y_{t-1}}} + \phi_{2_{y_{t-2}}} + \theta_{\varepsilon_{t-1}} + \varepsilon t, \qquad \varepsilon t \sim N(0,1)$$

Note that we have y_{t-1} and y_{t-2}: we wish to explain the dependent variable (y_t) by looking at the values of this variable in two past periods. This is consistent with the definition of AR.

Similarly, we have ε_{t-1}: we want to obtain some meaningful information contained in the lagged (one period in time) value of the residual. This is consistent with the definition of MA.

How to determine the order (p, d, q) of the ARIMA model?

- *For AR (p) and MA (q) components*, it usually suffices to examine the correlogram:



Autocorrelation Function Plot corresponds to MA, while Partial Autocorrelation Function Plot corresponds to AR.

We are interested in having a parsimonious model. This can be achieved by means of Wold Decomposition Theorem, which states that any invertible stationary model can be approximated with a linear model that is an infinite MA process. The practical method for achieving this is through backward substitution. Thus:

$$
\begin{aligned}
y_t &= \mu + \sum_{j=0}^{\infty} \psi_j a_{t-j}, \ \psi_0 = 1, \ \sum_{j=1}^{\infty} \psi_j^2 < \infty \\
y_t &= \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \psi_3 a_{t-3} + \cdots
\end{aligned}
$$

Following Wold Decomposition Theorem, we can opt for a parsimonious representation, whereby we model only the first spike, while the values of the successive ones can be obtained through backward substitution.

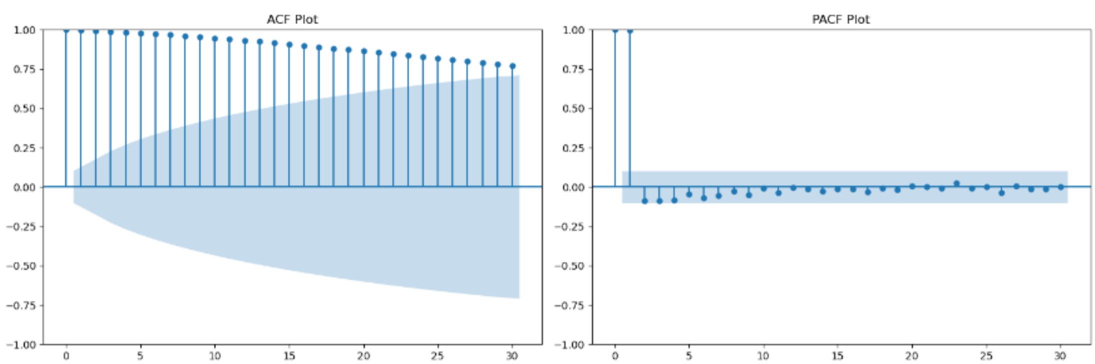
Numerical example:

$$y_t = 0.2 + 0.9 y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0,1),$$

We can model this process as AR(1). As has been indicated before, successive values could be obtained in an intuitive manner through backward substitution:

$$
\begin{aligned}
y_t &= 0.2 + {\color{red}0.9} y_{t-1} + \varepsilon_t \\
y_t &= 0.2 + {\color{red}0.9} \underbrace{(0.2 + {\color{red}0.9} y_{t-2} + \varepsilon_{t-1})}_{y_{t-1}} + \varepsilon_t \\
y_t &= 0.38 + 0.81 y_{t-2} + 0.9 \varepsilon_{t-1} + \varepsilon_t
\end{aligned}
$$

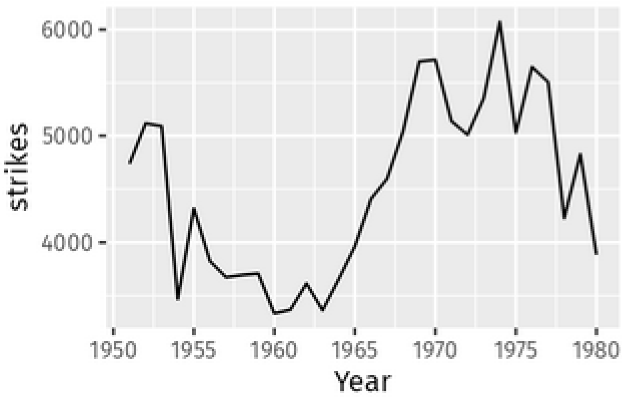
Returning to the previous correlogram:



We can determine the AR order to be 2 (p = 2), and the MA order to be 1 (q = 1).

- *For differencing order (d)*, we need to bear in mind that non-stationarity can be due to two motives: mean non-stationarity and / or variance non-stationarity.

Example (a non-stationary time series):

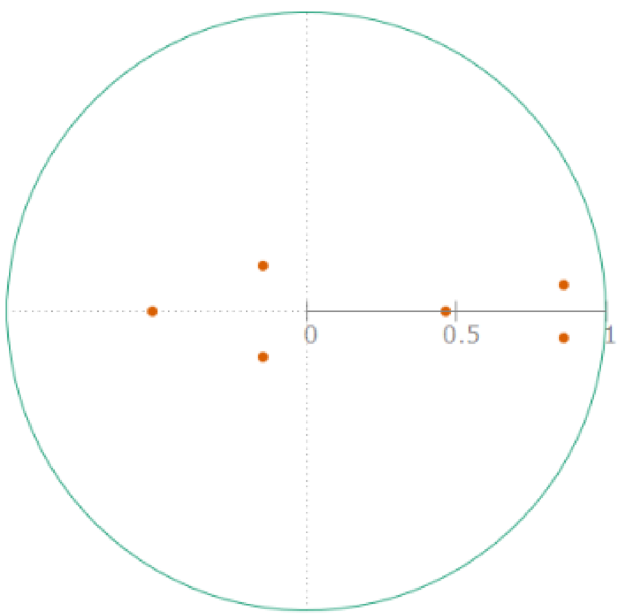


The statistical test is Augmented Dickey-Fuller (also known as unit root test), where:

H_0: A unit root is present in the time series.

H_1: A unit root is not present in the time series.

Why "unit root"? This has to do with the geometric representation of the inverse root of the polynomial:



In economics, the maximum differencing order to make time series stationary is usually 2. Thus:

Tcode	x	Definition
1	x	no transformation
2	Δx_t	First Difference
3	$\Delta^2 x_t$	Second Difference
4	$\log(x_t)$	Log
5	$\Delta \log(x_t)$	Log Difference
6	$\Delta^2 \log(x_t)$	Second Log Difference
7	$\Delta(x_t/x_{t-1} - 1.0)$	

Finally, for model selection, the model with the lowest AIC (Akaike Information Criterion) or BIC (Bayesian Information Criterion) is chosen.

To make sure that we are adequately capturing all the information, it is advisable to perform the Ljung-Box Test for residuals, where:

H_0: The data is not correlated.

H_1: The data is correlated.

Example of an ARIMA model output in Python:

```
Best ARIMA Model:
Order: (1, 1, 2), AIC: -838.65
                               SARIMAX Results
==============================================================================
Dep. Variable:      Unemployment Rate   No. Observations:                  315
Model:                 ARIMA(1, 1, 2)   Log Likelihood                 424.325
Date:                Fri, 15 Nov 2024   AIC                           -838.649
Time:                        20:24:34   BIC                           -819.902
Sample:                    01-31-1994   HQIC                          -831.158
                         - 03-31-2020
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
x1            -0.0112      0.017     -0.669      0.504      -0.044       0.022
ar.L1          0.8952      0.035     25.461      0.000       0.826       0.964
ma.L1         -0.7779      0.053    -14.813      0.000      -0.881      -0.675
ma.L2          0.2770      0.048      5.717      0.000       0.182       0.372
sigma2         0.0039      0.000     15.675      0.000       0.003       0.004
================================================================================
===
Ljung-Box (L1) (Q):                   0.03   Jarque-Bera (JB):
32.72
Prob(Q):                              0.87   Prob(JB):
0.00
Heteroskedasticity (H):               1.00   Skew:
```

```
0.39
Prob(H) (two-sided):                 0.99   Kurtosis:
4.38
================================================================================
===

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-
step).

Ljung-Box test for residuals:
      lb_stat  lb_pvalue
10  0.201057        1.0
```

2. Multivariate Time Series

The main motivation behind a multivariate representation is to obtain more information. Imagine that we wish to understand inflation. In a univariate representation we have only the past values of inflation and, possibly, the values of the residual. On the other hand, in a multivariate representation, we may include unemployment and economic growth TS – this is a rich source of additional information.

Modeling is usually performed by means of VAR (Vector AutoRegressive model).

Example: a VAR(3) with two variables:

$$\begin{bmatrix} y_{(1)t} \\ y_{(2)t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} \Phi_{11,1} & \Phi_{12,1} \\ \Phi_{21,1} & \Phi_{22,1} \end{bmatrix} \begin{bmatrix} y_{(1)t-1} \\ y_{(2)t-1} \end{bmatrix} + \begin{bmatrix} \Phi_{11,2} & \Phi_{12,2} \\ \Phi_{21,2} & \Phi_{22,2} \end{bmatrix} \begin{bmatrix} y_{(1)t-2} \\ y_{(2)t-2} \end{bmatrix} + \begin{bmatrix} \Phi_{11,3} & \Phi_{12,3} \\ \Phi_{21,3} & \Phi_{22,3} \end{bmatrix} \begin{bmatrix} y_{(1)t-3} \\ y_{(2)t-3} \end{bmatrix} + \begin{bmatrix} a_{(1)t} \\ a_{(2)t} \end{bmatrix}$$

Additionally, the VAR model may be extended to:

- SVAR (Structural VAR), in which we wish to account for propagation of shocks. The usual behaviour of the variables is captured in the dynamic matrix, while shocks are contained in the impact matrix:

$$\underbrace{\begin{bmatrix} x_t \\ y_t \end{bmatrix}}_{\mathbf{y}_t} = \underbrace{\overbrace{\begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}}^{\text{dynamic matrix}}}_{\mathbf{\Phi}} \underbrace{\begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix}}_{\mathbf{y}_{t-1}} + \underbrace{\overbrace{\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}}^{\text{impact matrix}}}_{\mathbf{B}} \underbrace{\begin{bmatrix} w_{xt} \\ w_{yt} \end{bmatrix}}_{\mathbf{w}_t}, \quad \underbrace{\begin{bmatrix} w_{xt} \\ w_{yt} \end{bmatrix}}_{\mathbf{w}_t} \sim \mathcal{N}\left(\underbrace{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}_{\mathbf{0}}, \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\mathbf{I}_2} \right)$$

- VECM (Vector Error Correction Model), which is used for cointegrated TS (time series that share a common unit root or stochastic trend):

- FAVAR (Factor Augmented VAR), in which PCA (Principal Component Analysis) is applied to estimate relevant features:

$$\begin{bmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{Nt} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} + \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \end{bmatrix} f_t + \begin{bmatrix} \nu_{1t} \\ \nu_{2t} \\ \vdots \\ \nu_{Nt} \end{bmatrix}$$

$$\mathbf{x}_t = \boldsymbol{\alpha} + \boldsymbol{\lambda} f_t + \boldsymbol{\nu}_t$$