MW82: Time Series Analysis, Tutorial II

Dr. Giulia Sabbadini

DICE

2023/24

Recap: Tutorial I

We went through:

- Introduction into Time Series
- Stationarity (informal)
- Trends and Seasonality

Today we will go through:

- Stationarity (more formal, ACF)
- Properties of the AR(1) model
- Estimate AR(1)

Recall: Covariance and correlation

Covariance measures joint variability of two random variables:

$$cov(X, Y) = E[(X - E[X])(Y - E[Y])]$$

Correlation: Standardized covariance (between -1 and 1)

$$\rho_{X,Y} = \frac{cov(X, Y)}{\sigma_X \sigma_Y}$$

where σ is the standard deviation.

Formal: Stationarity

- **Strict** stationarity: A stationary process is identically distributed across time
- (Weak) stationarity, covariance-stationarity:
  A time series $\{y_t\}_{t=1}^{T}$ is stationary iff
  1. $\mu_t = \mu_{t-s} = \mu < \infty$ for all t, s
  2. $\mathrm{var}(y_t) = \mathrm{var}(y_{t-s}) = \sigma^2 < \infty$ for all t, s
  3. $\mathrm{cov}(y_t, y_{t-s}) = \mathrm{cov}(y_{t-j}, y_{t-j-s}) = \gamma(s)$ for all t, j, s

  1.: mean function is constant and finite

  2.: variance is constant and finite

  3.: covariance depends only on the time distance s between the two elements of the time series, but not on time t itself

Why is stationarity important?

- If a time-series exhibits “similar” behavior, one can then proceed with the modeling efforts.
- **Wold’s theorem**: Any covariance-stationary time series can be arbitrarily well approximated by an ARMA-Model
- **A**uto**r**egressive-**M**oving-**A**verage models: combination of autoregressive (AR) & moving average (MA) parts

ARMA models

- Autoregressive process of lag order p, AR(p), is described as:

$$y_t = c + \sum_{i=1}^{p} \phi_i \, y_{t-i} + \varepsilon_t$$

- Moving-average process of order q, MA(q), is:

$$y_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \, \varepsilon_{t-i}$$

- ARMA(p,q):

$$y_t = \varepsilon_t + \sum_{i=1}^{p} \phi_i \, y_{t-i} + \sum_{i=1}^{q} \theta_i \, \varepsilon_{t-i}$$

ε_t is white noise.

Autocorrelation function (ACF)

The ACF of $\{y_t\}_{t=1}^{T}$ gives correlations between y_t and y_{t-s} for $s = 0, 1, 2, \ldots$.

$$\frac{cov(y_t, y_{t-s})}{sd(y_t)\ sd(y_{t-s})} = \frac{cov(y_t, y_{t-s})}{Var(y_t)}$$

- Useful to find non-stationarities
- To assess the randomness and stationarity of a time series
- To determine whether trends and seasonal patterns are present.

AR(1) Model

$$y_t = c + \phi y_{t-1} + \varepsilon_t$$

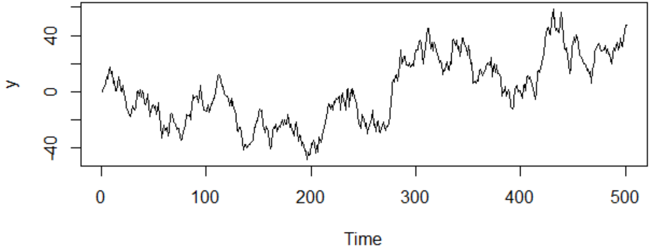
where ε_t is white noise (serially uncorrelated RV with mean 0 and finite variance σ^2).
Different values of parameter ϕ imply:

- $|\phi| < 1$: stationary
- $\phi = 1$: unit root (non-stationary)
- $|\phi| > 1$: explosive (non-stationary)

Unit root ($\phi = 1$)

Since $\phi = 1$, y_t is the sum of historical inputs, i.e. shocks do not vanish.

$$
\begin{aligned}
y_1 &= \varepsilon_1 \\
y_2 &= 1 \cdot y_1 + \varepsilon_2 = \varepsilon_1 + \varepsilon_2 \\
y_3 &= 1 \cdot y_2 + \varepsilon_3 = \varepsilon_1 + \varepsilon_2 + \varepsilon_3
\end{aligned}
$$



Testing for unit roots is very important (later in the course)!

Properties of an AR(1) model

For any stationary AR(1) process we have:

- Mean:

$$E[y_t] = \mu = \frac{c}{1-\phi}$$

- Variance:

$$Var(y_t) = \frac{\sigma^2}{1-\phi^2}$$

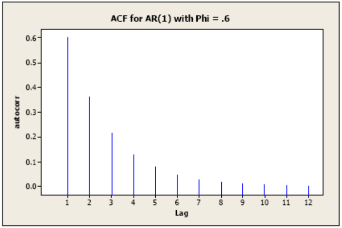
- Autocorrelation (correlation between observations s time periods apart) is:

$$\rho_s = \phi^s$$

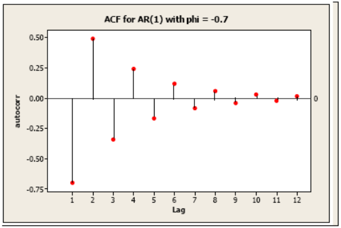
- Autocorrelation function (ACF) is the collection of these values (ρ_s) for lags $s = 1, 2, 3, \ldots$

ACF Patterns for AR(1) Model

- positive value of ϕ: exponential decrease (tapering) to 0 as lag s increases (a)
- negative value of ϕ: exponential decrease (tapering, alternating signs) to 0 as lag s increases (b)



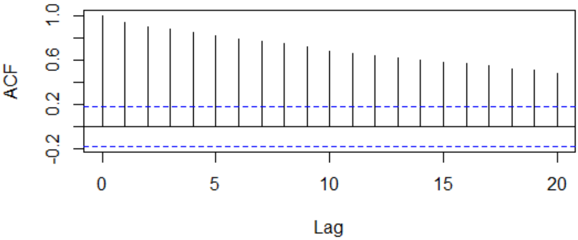
(a)



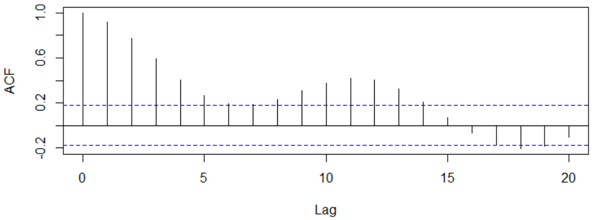
(b)

AR(1): Non-stationarities in ACF

- How does a trend look like?



- How does seasonality look like?



Useful commands

Lag plot: `lag.plot` or `astsa::lag1.plot`

Get coefficients from regression: `coef()`

forecast package: many useful commands (like `Acf`).

Exercise I

Dataset: Annual number of earthquakes (> 7.0) for 99 years from 1900.

1. Load the *quakes.dat* dataset in R (Hint: use the *scan* function).
2. Create a time series plot and add a mean line. Does the series look stationary?
3. Plot the ACF of this time series. Do you find evidence for non-stationarity?
4. Create a lag-1-plot (y_t, y_{t-1}). Is there a relationship between those values?
5. Estimate an AR(1) model ($y_t = c + \phi y_{t-1} + \varepsilon$) with OLS and report your parameter estimates.
6. How does a one-unit-shock in t affect the number of earthquakes 3 years later.