Time Series Econometrics

Fabio Bacchini, Stefano Tonellato

Eurostat Training Course
Venice, 11-13 February 2020

GOPA GOPA

The contractor is acting under a specific contract concluded with the Commission

Introduction of teachers

Name	Fabio Bacchini (Course leader)
Function	Head Division of data analysis and research in economics, environment and social analysis - Istat
Email	bacchini@istat.it

Name	Stefano tonellato
Function	Associate professor of statistics Ca' Foscari University
Email	stone@unive.it.it

This presentation includes previous contribution by Roberto Iannaccone (Istat-Italy) and Umberto Triacca (University of L'Aquila-Italy)

Introduction of the participants

- Where are you from?
- What is your experience with time series?
- What is your experience with statistical or econometrics software?
- What is your experience with language R?

Outlines

- The goal of the course is to provide participants with a basic knowledge of modern time series econometrics
- Attention will focus on univariate and multivariate approach
- Participants will be introduced to the topics through several exercises developed in the R environments

Expected results

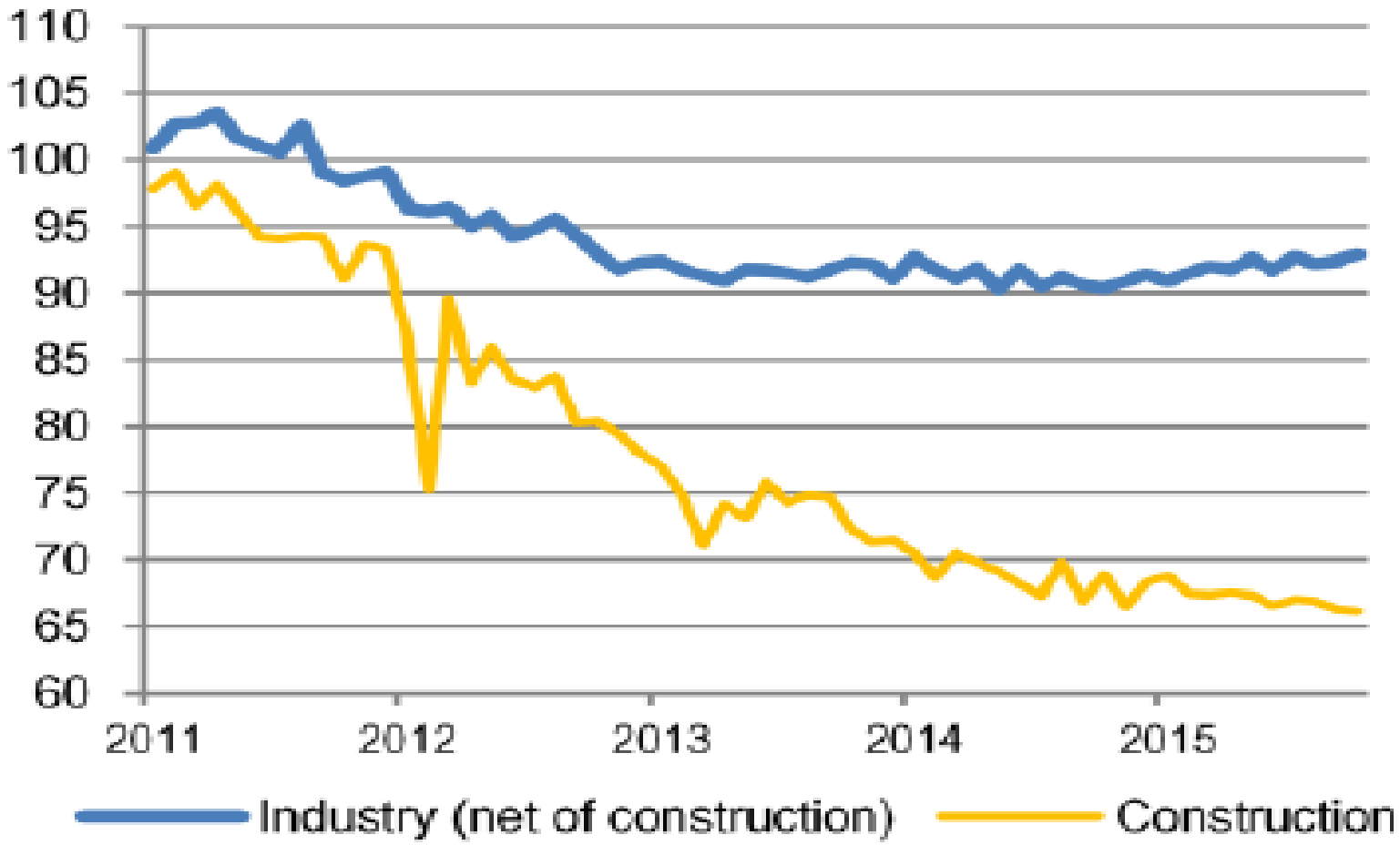
After this training activity, participants will be expected to perform economic analysis on their domain of interest using univariate or multivariate models

Main topics

- Introduction to time series and to R environment
- Univariate time series: AR, MA, ARIMA process (definition, estimation and practical exercises)
- Forecasting with time series: introduction, nowcasting, type of forecasting, practical exercise
- Revision analysis and vintages policy: the approaches followed by international institutions
- Multivariate time series: introduction to VAR
- Multivariate time series: Cointegration and VECM

Univariate

Italian production indices - 2010=100, **seas. adj.**

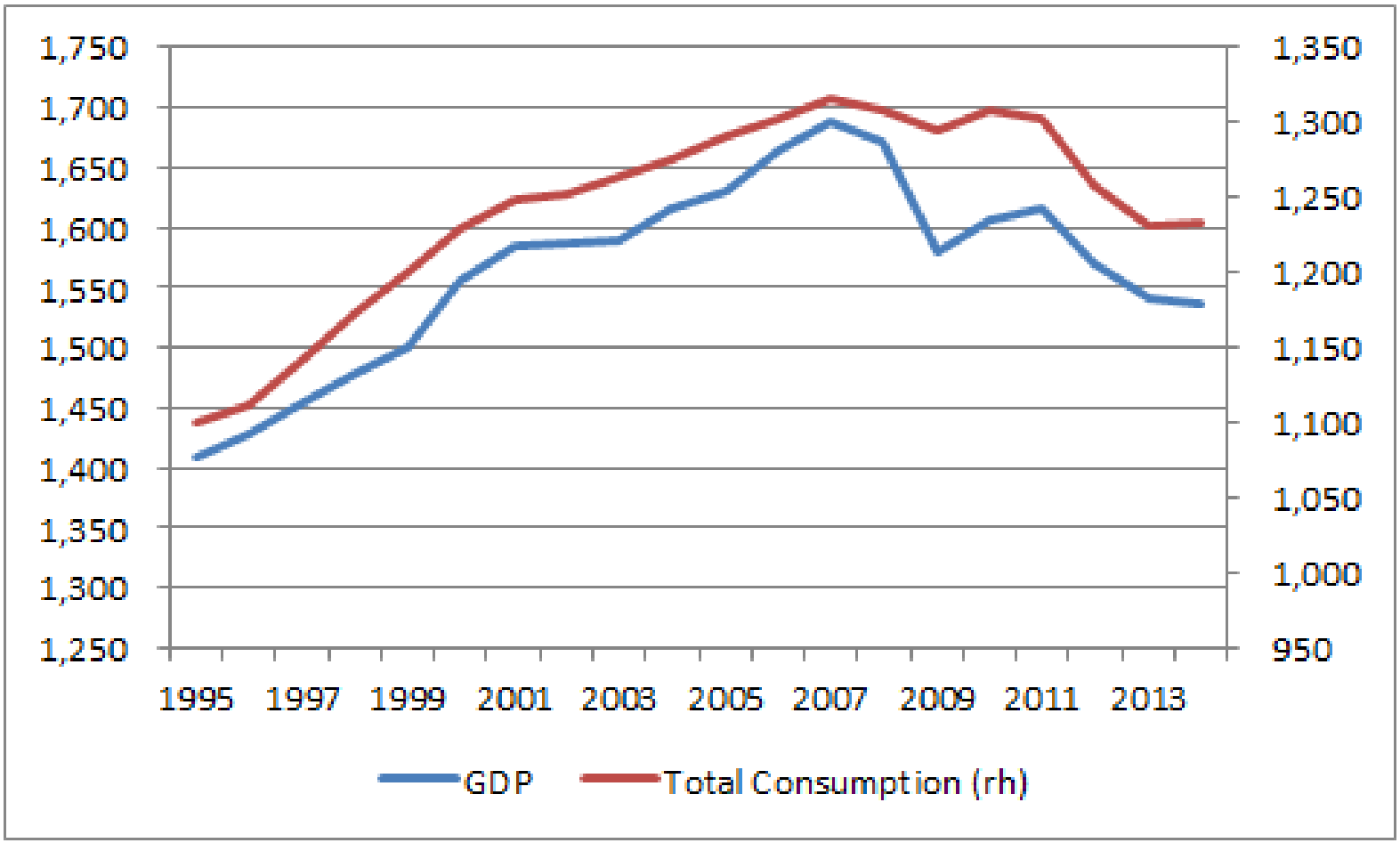


GOPA GOPA

European Commission

Multivariate

Italian GDP and Consumption



Statistical model

es. (S)AR(I)MA

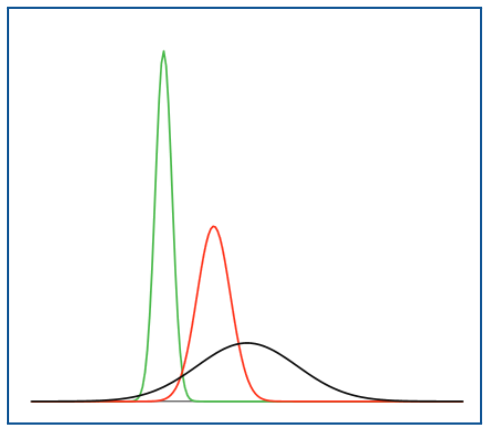
Observed time series

es. Industrial production

es. GDP, Consumption

Traditional statistical framework

We perform an experiment, the outcome of each experiment could be represented by a **Random variable** X_i. A random variable is represented by its probability density function $f(X)$. For example the normal distribution



$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Normal distribution is characterized by 2 parameters and the distribution. We can extend this approach to different variables $i = 1, 2, ..., n$

Time series

is characterized by different observation along the time (month, quarter, year). We suppose that each observation is an outcome of a random variable. In such way we introduce the concept of a stochastic process

$$x = \{x_t, t \in Z\}$$

where Z is a set of integer numbers. For example we can think to the annual GDP observed from 1990 to 2014 as a particular realisation of a stochastic process $x_{t1} \dots x_{tn}$ where $t1 = 1990$...
In general a stochastic process is described by its **probability distribution** $f(x_1, x_2, \dots, x_n)$ but we can 'weakly' define it by its **moments**.

GOPA GOPA

European Commission

Moments:

n means $E(x_1), E(x_2), \ldots, E(x_n)$

n variances $V(x_1), V(x_2), \ldots, V(x_n)$

and n(n-1)/2 covariances $Cov(x_i, x_j), i < j$

Stationarity - 1

Stationarity requires that the process is in a particularly state of equilibrium, that is when its statistical properties are non dependent from the time (Mills, pag. 64)

- **strong stationarity**: when the join probability distribution at any set of times $t_1, t_2, \ldots, t_m$ must be the same as the joint probability distribution at time $t_1 + k, t_2 + k, \ldots, t_m + k$ where k is an arbitrary shift along the time series. For $m = 1$ this implies that the marginal probability distribution at time t is the same ad the marginal probability distribution at any other point in time: $f(x_t) = f(x_{t+k})$

Hence the marginal distribution does not depend on time, which in turn implies that the mean and the variance must be constant

$$E(x_1) = E(x_2) = \ldots = E(x_n) = E(x_t) = \mu$$

$$V(x_1) = V(x_2) = \ldots = V(x_n) = V(x_t) = \sigma^2$$

Stationarity - 2

If $m = 2$ stationarity implies that all bivariate distribution f(x_t, x_{t+k}) do not depend on t and the covariance are function only of the lag k, and not of time t for all k:

$$Cov(x_1, x_{1+k}) = Cov(x_2, x_{2+k}) = \ldots = Cov(x_n, x_{n+k})$$

- the stationarity assumption implies that the mean and the variance of the process are constant and that the autocovariance and autocorrelation depend only on the lag k

Weak stationarity: this property involves only the first (mean) and the second moment (variance and autocovariance) but does not refer to the probability distribution.

$$E(x_t) = \mu$$
$$Var(x_t) = \gamma_0$$
$$Cov(x_t, x_{t-k}) = \gamma_k.$$

GOPA GOPA

European Commission

Ergodicity

- T. Mills (Time series techniques for economists) 'it should also be noted that procedure of using a single realisation to infer the unknown parameters of the joint probability distribution is only strict valid if the process is **ergodic**, which roughly means that the sample moments for the finite stretches of the realisation approach their population moments as the length of the realisation becomes infinite'
- an ergodic process is characterized by low memory when n, the length of the time series, is high

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} Cov(x_n, x_{n-k}) = 0$$

We shall state the conditions that guarantee the ergodicity of the stochastic processes under investigation.

Sample moments

- mean: $\hat{\mu} = \frac{\sum_{t=1}^{T} x_t}{T}$
- variance: $\hat{\gamma}_0 = \frac{\sum_{t=1}^{T} (x_t - \hat{\mu})^2}{T}$
- autocovariance $\hat{\gamma}_k = \frac{\sum_{t=k}^{T} (x_t - \hat{\mu})(x_{t-k} - \hat{\mu})}{T}$
- autocorrelation: $\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$

when the process is stationary and ergodic this sample moments are consistent estimators for the moments in the population

A time series could be characterised by some deterministic component that can be perfectly predicted from past values of itself, such as:

- costant mean
- trend (linear, exponential, quadratic)
- periodic component

(see for example Diebold, Econometrics, p.127)

Wold's decomposition (Wold, 1938)

Every weakly stationarity, ergodic, stochastic process (x_t) can be written as a linear combination (or linear filter) of a sequence of uncorrelated random variables, where μ ia a general representation of a deterministic component

$$x_t = \mu + \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \dots$$

$$x_t = \mu + \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} \qquad \psi_0 = 1$$

$$E(\epsilon_t) = 0$$

$$Var(\epsilon_t) = \sigma^2$$

$$Cov(\epsilon_t, \epsilon_s) = 0 \qquad t \neq s.$$

$$\sum_{j=0}^{\infty} |\psi_j|^2 < \infty$$

ϵ_t is the building block for many statistical models, the white noise process

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

If ϵ_t is also normally distributed, then it follows that ϵ_t and ϵ_s are also independent for $t \neq s$, and we say that ϵ_t is normal white noise, or Gaussian white noise.

$$\epsilon_t \sim iidN(0, \sigma^2)$$

GOPA GOPA

European Commission

Unconditional and conditional moments

"Another characterization of dynamics involves the mean and variance of a process, conditional upon its past. In particular, we often gain insight into the dynamics in a process by examining its conditional mean" (Diebold, Ecoonmetrics, p.169)

As in introduction to conditional and unconditional moments we refer to $\epsilon_t \sim iidN(0, \sigma^2)$. As noted before the unconditional mean of ϵ_t is zero and the unconditional variance is σ^2

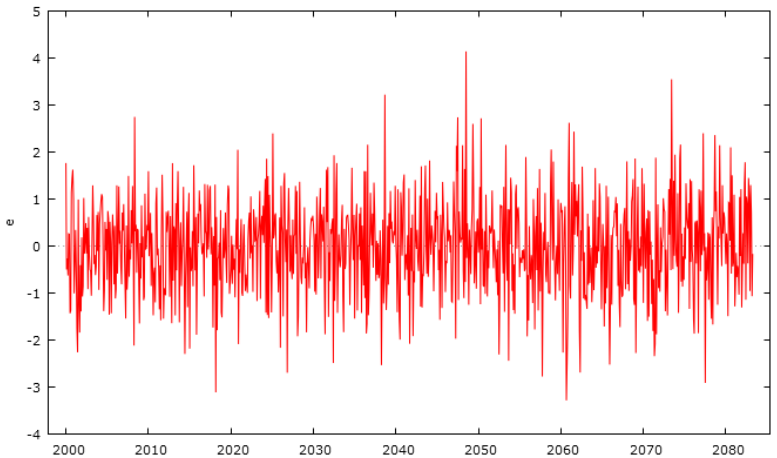
- consider the conditional mean and variance according to the information set Ω_{t-1}
- the conditions contains either the past history of the observed series $\Omega_{t-1} = y_{t-1}, y_{t-2}, \dots$
- or the past history of the shocks $\Omega_{t-1} = \epsilon_{t-1}, \epsilon_{t-2}, \dots$
- **they are the same in the white noise case**

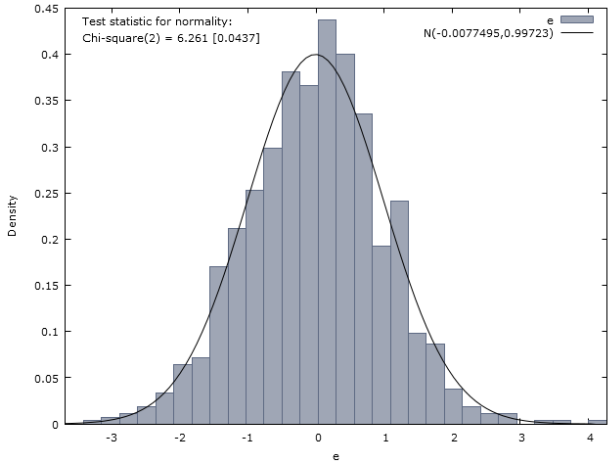
In the simple case of WN:

$$E(y_t|\Omega_{t-1}) = 0$$

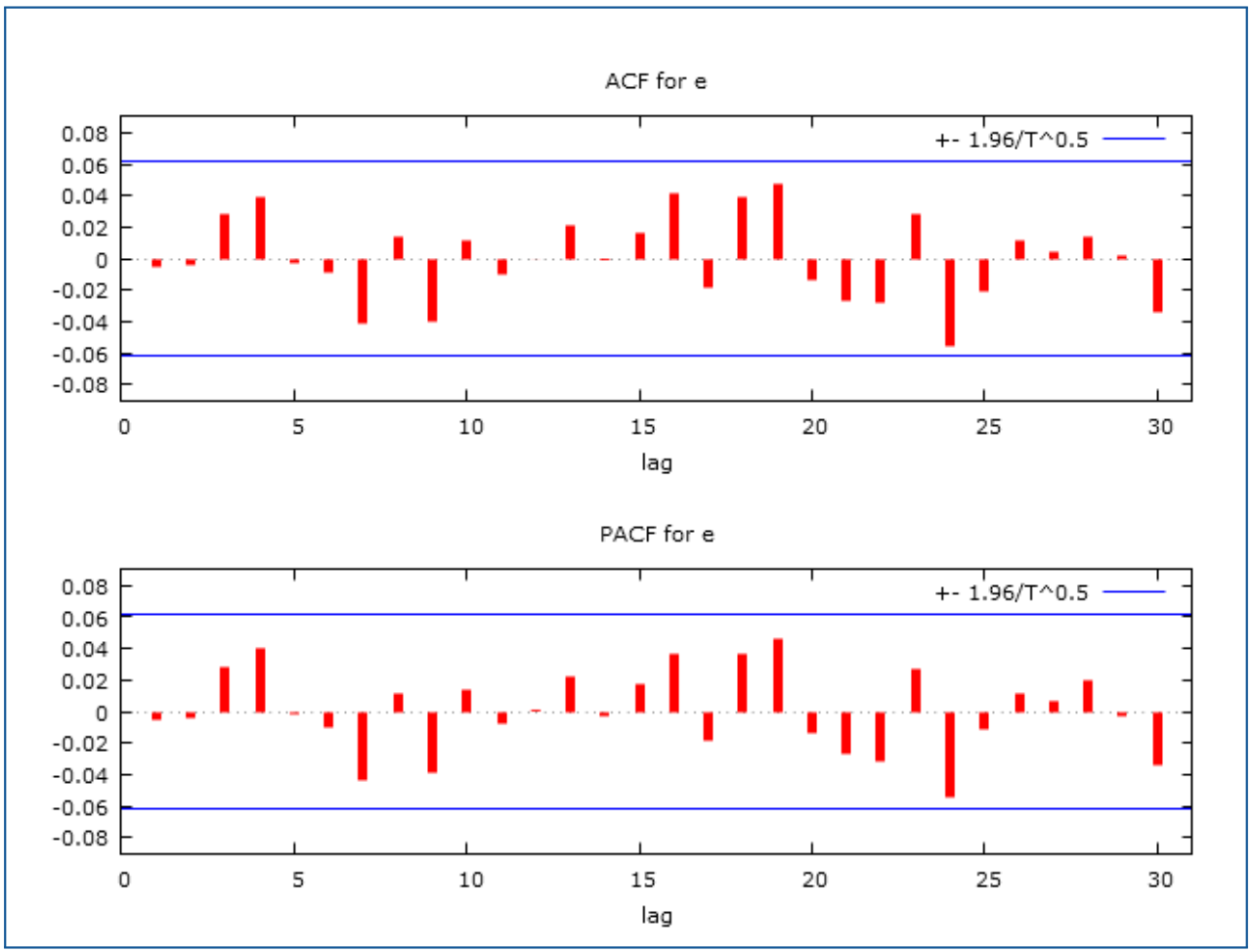
$$var(y_t|\Omega_{t-1}) = E[(y_t - E(y_t|\Omega_{t-1}))^2] = \sigma^2$$

White noise (1000 obs.)

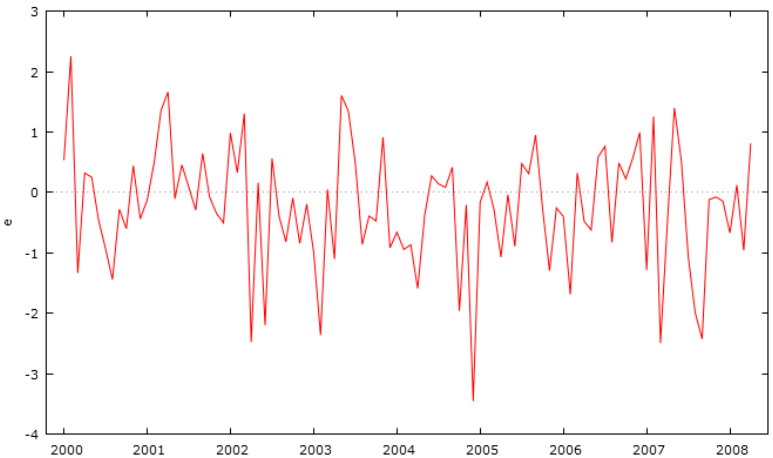


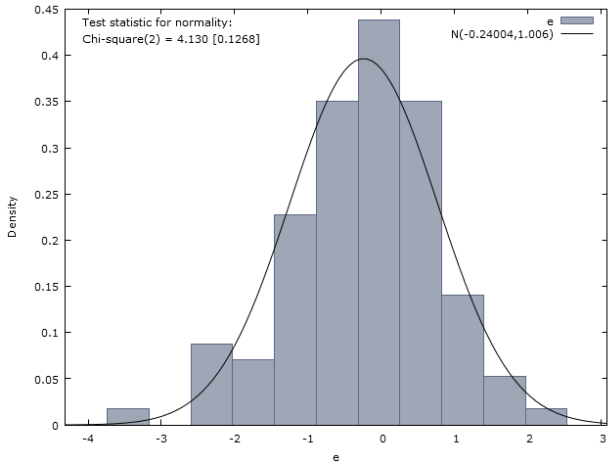


White noise (1000 obs.)



White noise (100 obs.)





MA(1)

- from Wold's theorem $\psi_1 = \theta$ and $\psi_j = 0$, for $j \leq 2$
- $x_t = \epsilon_t + \theta_1 \epsilon_{t-1}$
- $x_t = (1 + \theta_1 L)\epsilon_t$

where L is lag operator:

$$Lx_t = x_{t-1}$$

and

$$L^n x_t = x_{t-n} \text{ and } L^0 x_t = x_t$$

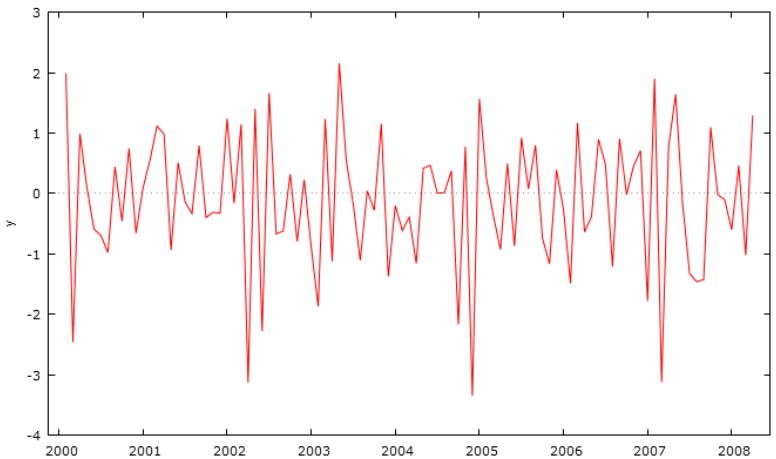
and

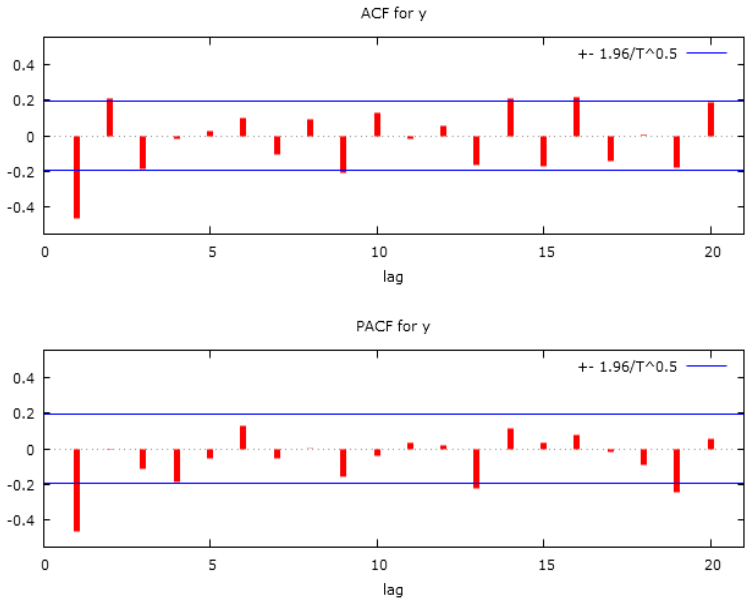
$$\Delta = (1 - L)$$

GOPA GOPA

European Commission

MA(1), *series* $x_t = \epsilon_t - 0.5 * \epsilon_{t-1}$





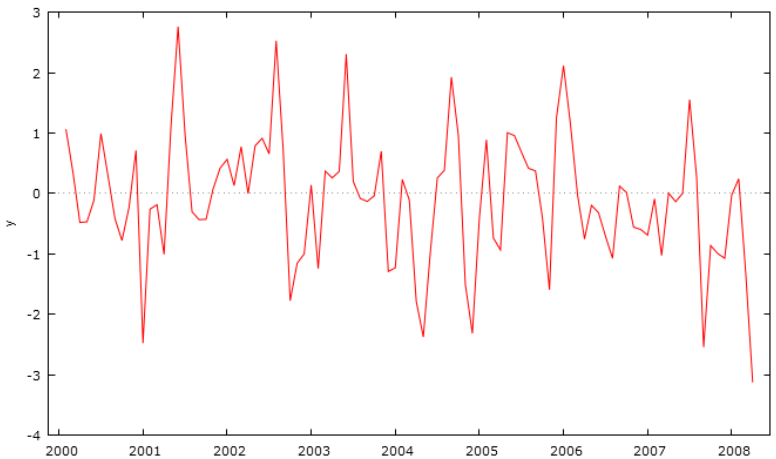
Note that

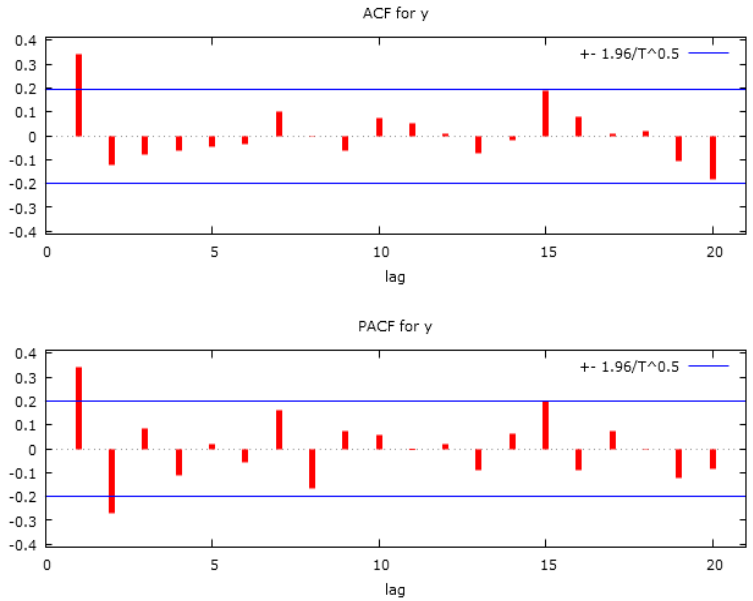
- in general the correlation between two random variables is often due to both variables being correlated with a third variable
- in the context of time series a large portion of the correlation between x_t and x_{t-k} can be due to the correlation these variables with the other lags $x_{t-1}, x_{t-2} \ldots$
- the lag k partial autocorrelation is the partial regression coefficient ϕ_{kk} in the k order autoregression

$$x_t = \phi_{k1} x_{t-1} + \phi_{k2} x_{t-2} + \ldots + \phi_{kk} x_{t-k} + a_t$$

for example for autoregressive process a PACF is zero for lags larger than p

MA(1), *series* $x_t = \epsilon_t + 0.5\epsilon_{t-1}$

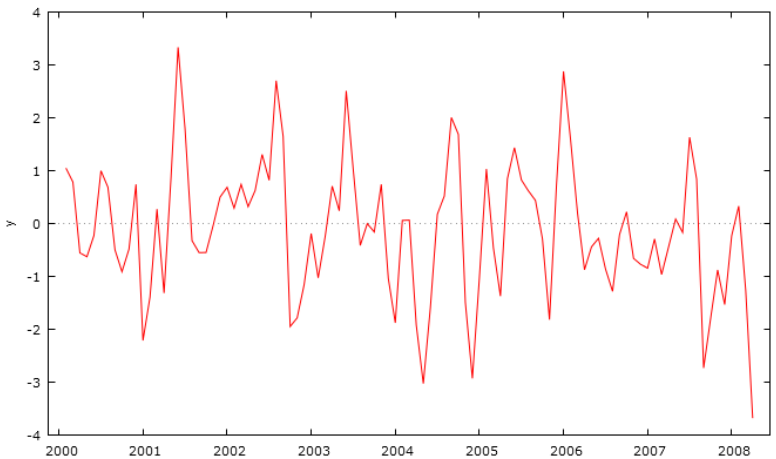


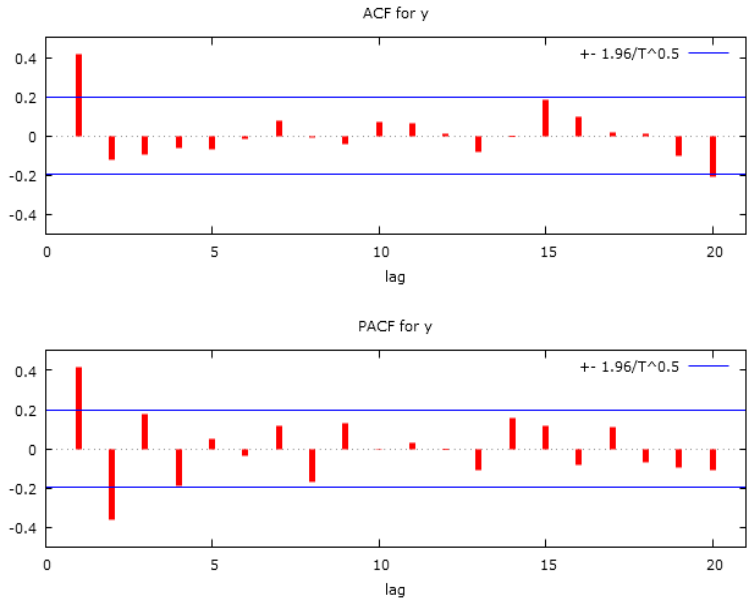


GOPA GOPA

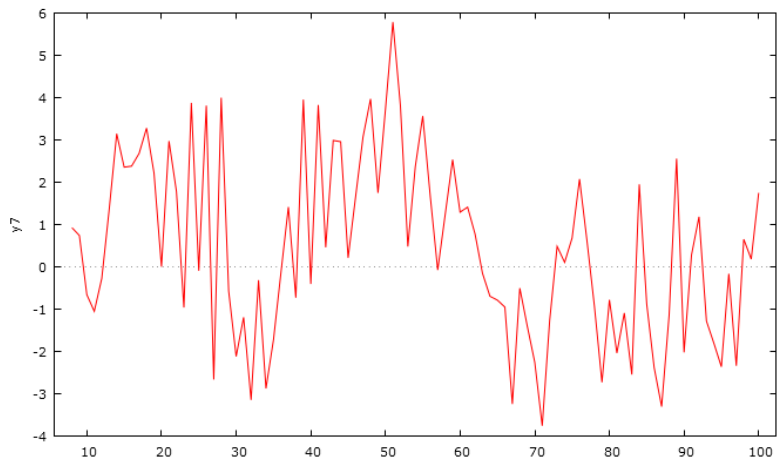
European Commission

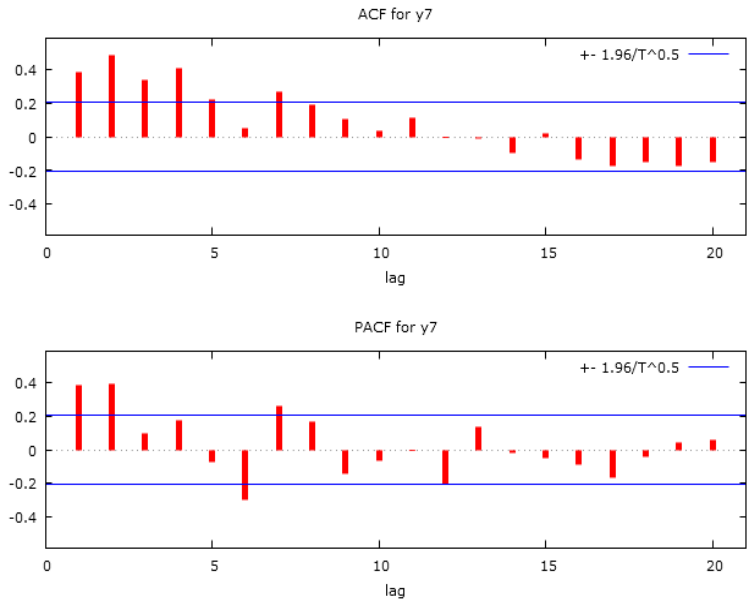
MA(1), *series* $x_t = \epsilon_t + 0.9\epsilon_{t-1}$





MA(7) *series* $x_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_7 \epsilon_{t-7}$





Note MA process could generate persistence but we need for more terms

Example MA(1)

$$x_t = \epsilon_t + \theta\epsilon_{t-1}$$

Mean=0, Variance:

$$E(x_t^2) = E(\epsilon_t + \theta\epsilon_{t-1})^2 = E(\epsilon_t^2) + \theta^2 E(\epsilon_{t-1}^2) + 2\theta E(\epsilon_t\epsilon_{t-1}) = (1+\theta^2)\sigma^2$$

autocovariance of order 1

$$E(x_t x_{t-1}) = E[(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_{t-1} + \theta\epsilon_{t-2})$$

$$E(\epsilon_t\epsilon_{t-1}) + \theta E(\epsilon_t\epsilon_{t-2}) + \theta E(\epsilon_{t-1}^2) + \theta^2 E(\epsilon_{t-1}\epsilon_{t-2}) = \theta\sigma^2$$

autocorrelation

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\theta}{1+\theta^2}$$

Moments of a general MA(q) process

$$x_t = \sum_{i=0}^{q} \theta_i \epsilon_{t-i}$$

where $\theta_0 = 1$;

$$E(x_t) = E\left[\sum_{i=0}^{q} \theta_i \epsilon_{t-i}\right] = \sum_{i=0}^{q} \theta_i E(\epsilon_{t-i}) = 0$$

$$V(x_t) = E\left[\sum_{i=0}^{q} \theta_i \epsilon_{t-i}\right]^2$$

but

$$\left(\sum_{i=0}^{q} \theta_i \epsilon_{t-i}\right)^2 = \sum_{i=0}^{q} \theta_i^2 \epsilon_{t-i}^2 + \sum_{i=0}^{q} \sum_{i \neq j} \theta_i \theta_j \epsilon_{t-i} \epsilon_{t-j}$$

$$E[(x_t^2)] = E[\sum_{i=0}^{q} \theta_i^2 \epsilon_{t-i}^2] = \sum_{i=0}^{q} \theta_i^2 \sigma^2 = \sigma^2 \sum_{i=0}^{q} \theta_i^2$$

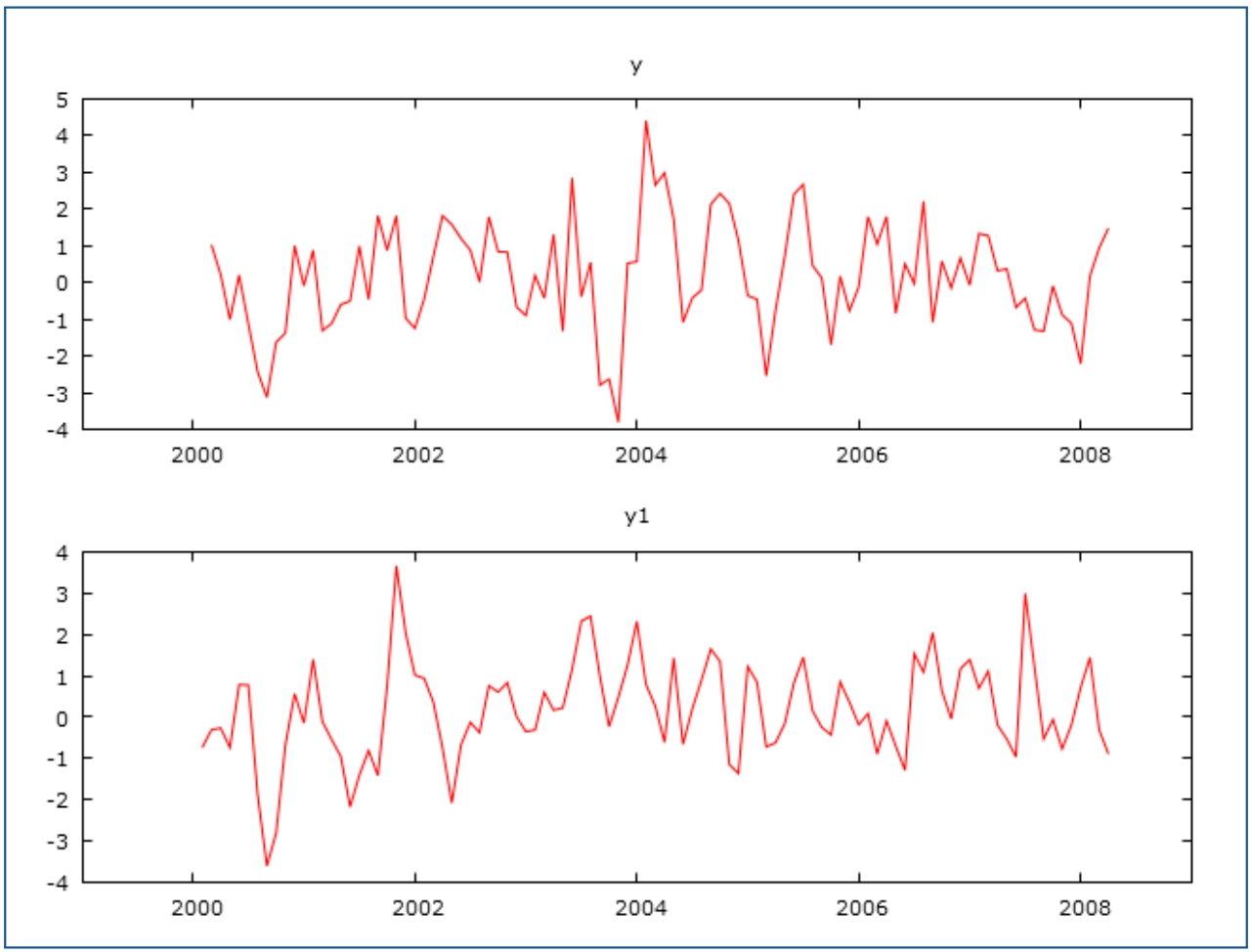
Similarly, it can be shown that

$$\gamma_k = E(x_t, x_{t+k}) = \sigma^2 \sum_{i=0}^{q} \theta_i \theta_{i+k}$$

MA characteristics

- an MA process is a linear combination of white noise
- for k>q autocovariance are equal to zero
- persistence is related to its order
- **I suffer from short term memory loss, Dory says to Marlin. It's true, I forget things almost instantly. It runs in my family. Well, at least I think it does, Dory says.**

compare MA2 (y, 0.5, 0.8) vs. MA(1) (y1, 0.5)



AR(1) characteristics:

$$x_t = \phi x_{t-1} + \epsilon_t, \qquad \epsilon_t \sim WN(0, \sigma^2)$$
$$x_t - \phi x_{t-1} = \epsilon_t$$
$$(1 - \phi L) x_t = \epsilon_t$$

that implies

$$\begin{aligned} x_t &= (1 - \phi L)^{-1} \epsilon_t \\ &= (1 + \phi L + \phi^2 L^2 + \ldots) \epsilon_t \\ &= \epsilon_t + \phi \epsilon_{t-1} + \phi^2 \epsilon_{t-2} + \ldots \end{aligned}$$

Clearly this linear representation ($\psi_j = \phi^j$) 'will converge as long as $|\phi| < 1$ which is therefore the stationarity condition

Autocovariance: multiply by x_{t-k}, $k \geq 0$ and take expectation:

$$E(x_t x_{t-k}) - \phi E(x_{t-1} x_{t-k}) = E(\epsilon_t x_{t-k})$$

or

$$\gamma_k - \phi \gamma_{k-1} = E(\epsilon_T x_{t-k})$$

$$E(\epsilon_t x_t) = E[\epsilon_t(\epsilon_t + \phi\epsilon_{t-1} + \ldots)] = \sigma^2 \qquad k = 0$$

$$E(\epsilon_t x_t) = 0 \qquad k > 0$$

$$\gamma_0 - \phi\gamma_1 = \sigma^2$$
$$\gamma_k - \phi\gamma_{k-1} = 0 \quad k \geq 1$$

Since $\gamma_1 = \phi\gamma_0$

$$\gamma_0 = \frac{\sigma^2}{1-\phi^2}$$

and

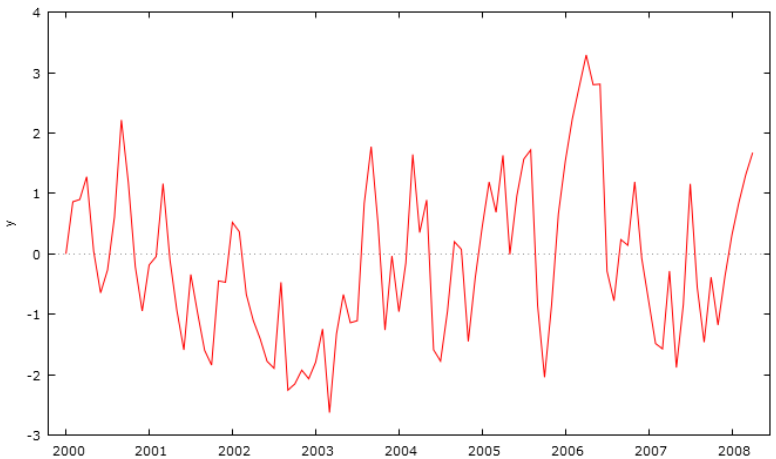
$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi\rho_{k-1}$$

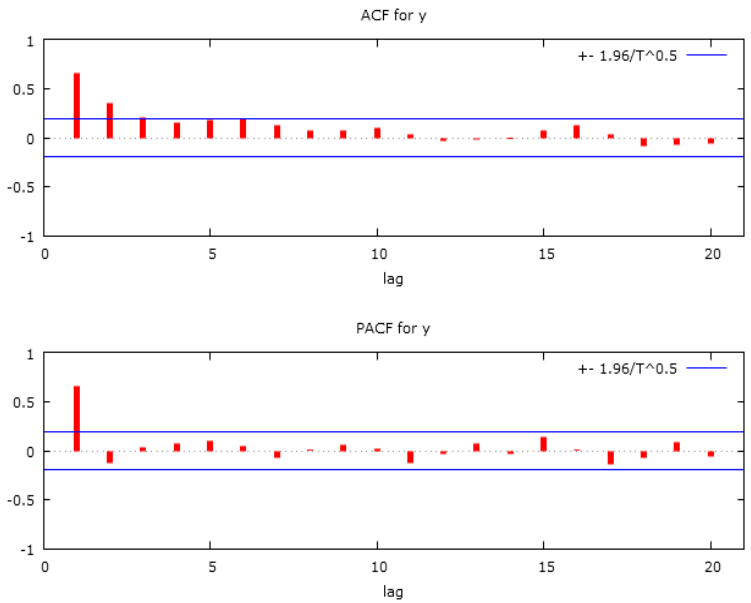
which, on solving, gives:

$$\rho_k = \phi\rho_{k-1} = \phi^2\rho_{k-1} = \ldots = \phi^k\rho_0 = \phi^k$$

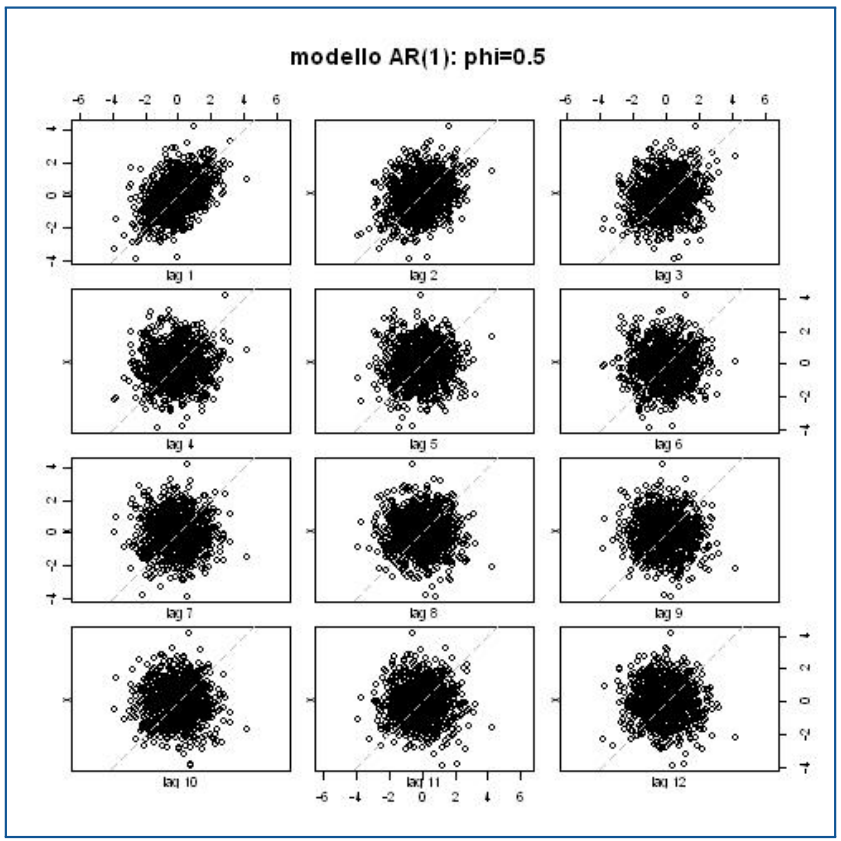
for $k = 0, 1, 2, \ldots$. Thus if $|\phi| < 1$ the autocorrelation function decays exponentially fast to zero

AR(1) low persistence (positive parameter $\phi = 0.5$)

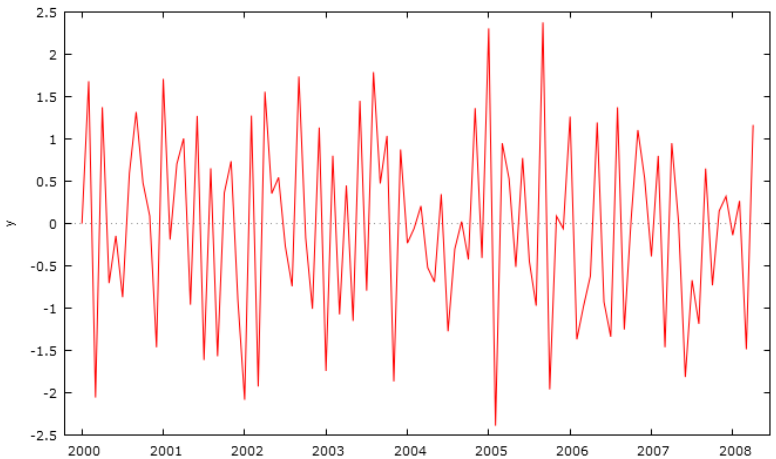


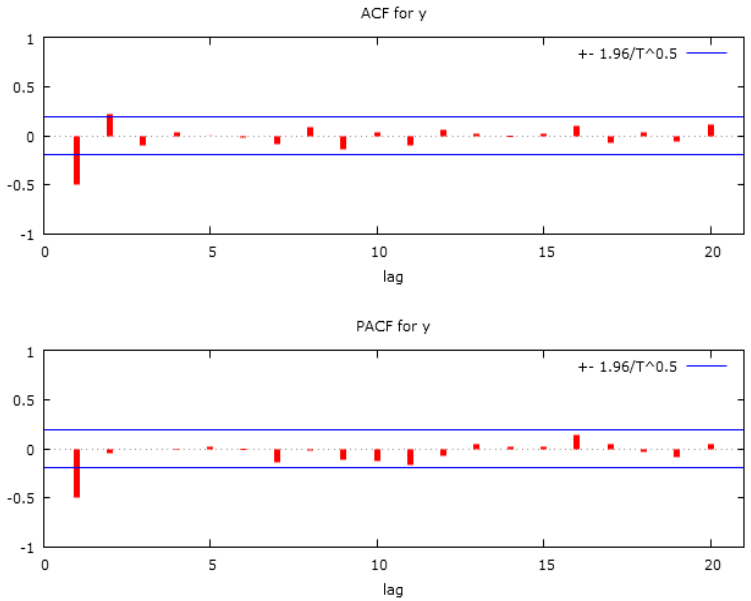


AR(1) low persistence

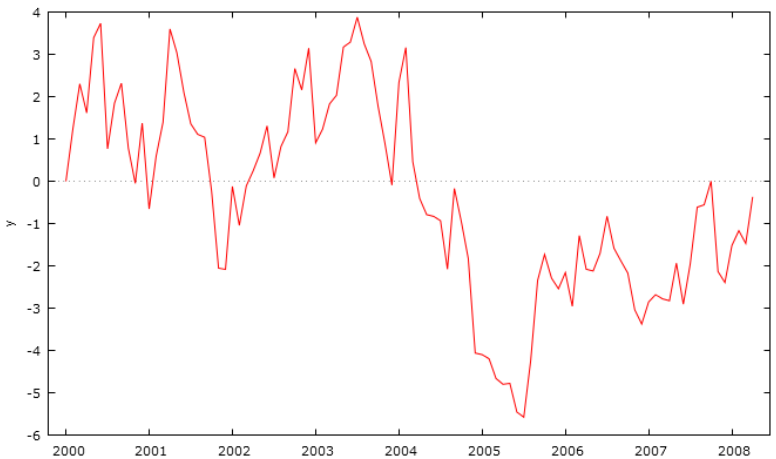


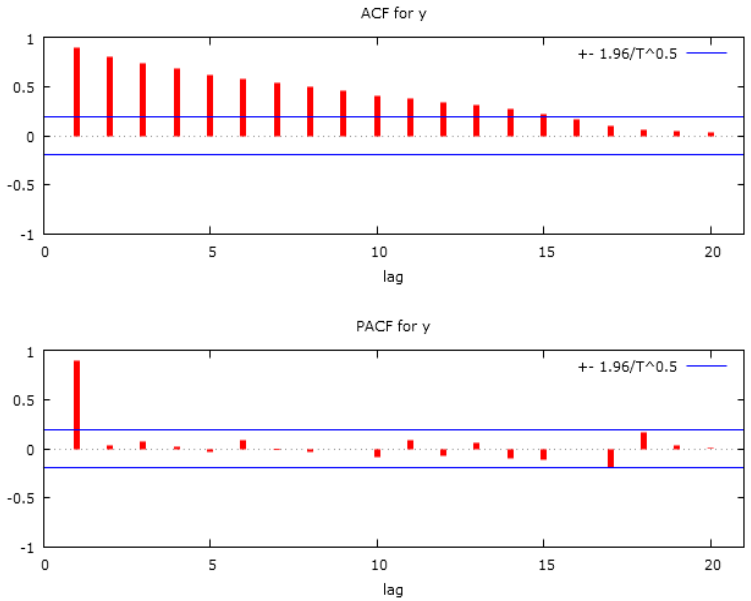
AR(1) low persistence (negative parameter $\phi = -0.5$)



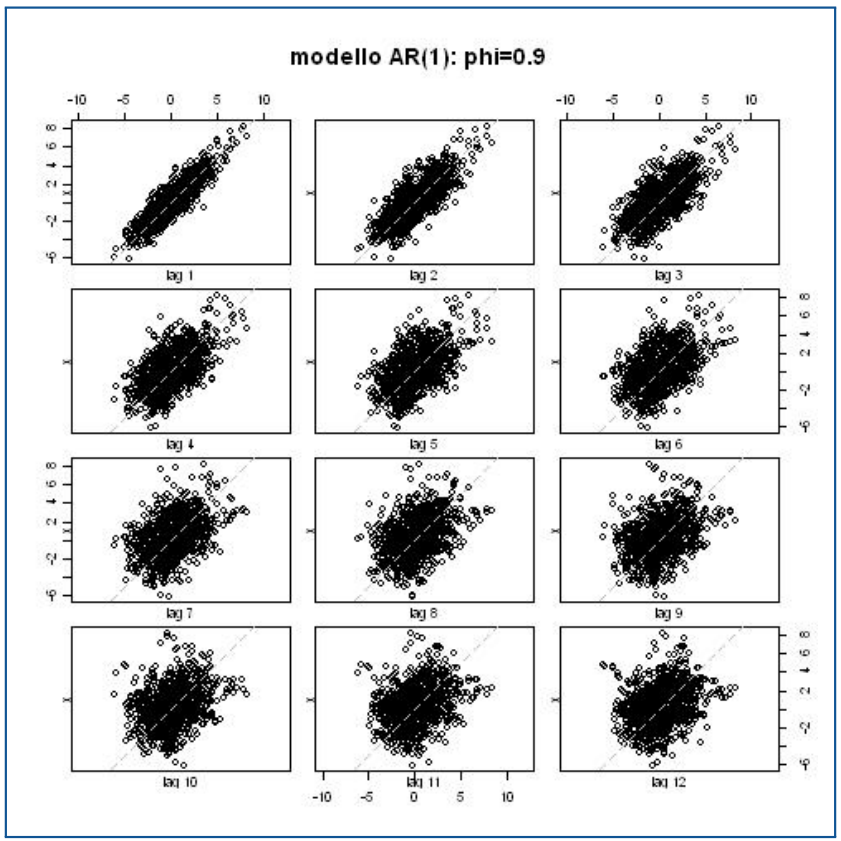


AR(1) high persistence (positive parameter $\phi = 0.9$)





AR(1) high persistence



GOPA GOPA

European Commission

Autoregressive Moving Average (ARMA) models (Box and Jenkins,1976)

$x_t \sim ARMA(p, q)$

$$\begin{aligned} x_t - \mu &= \phi_1(x_{t-1} - \mu) + \cdots + \phi_p(x_{t-p} - \mu) + \\ &\quad \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q} \\ \varepsilon_t &\sim WN(0, \sigma^2) \end{aligned}$$

Let $\mu = 0$ with no loss of generality

$$\begin{aligned}(1 - \phi_1 B - \cdots - \phi_p B^p) x_t &= (1 + \theta_1 B + \cdots + \theta_p B^p) \varepsilon_t \\ x_t - \phi_1 x_{t-1} - \cdots - \phi_p x_{t-p} &= \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}\end{aligned}$$

GOPA GOPA

European Commission

$$
\begin{aligned}
\phi(B)x_t &= \theta(B)\varepsilon_t \\
\varepsilon_t &= WN(0, \sigma^2) \\
\phi(B) &= 1 - \phi_1 B - \cdots - \phi_p B^p \\
\theta(B) &= 1 + \theta_1 B + \cdots + \theta_q B^q
\end{aligned}
$$

Consider the equation

$$a_0 + a_1 z + a_2 z^2 + a_n z^n = 0$$

and suppose there exists a z^* such that

$$a_0 + a_1 z^* + a_2 {z^*}^2 + a_n {z^*}^n = 0.$$

Then z^* is a **root** of the polynomial

$$p(z) = a_0 + a_1 z + a_2 z^2 + a_n z^n.$$

Stationarity and invertibility

$\{x_t\}$ is stationary if

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z \neq 0$$

for all $|z| \leq 1$. x_t is invertible if

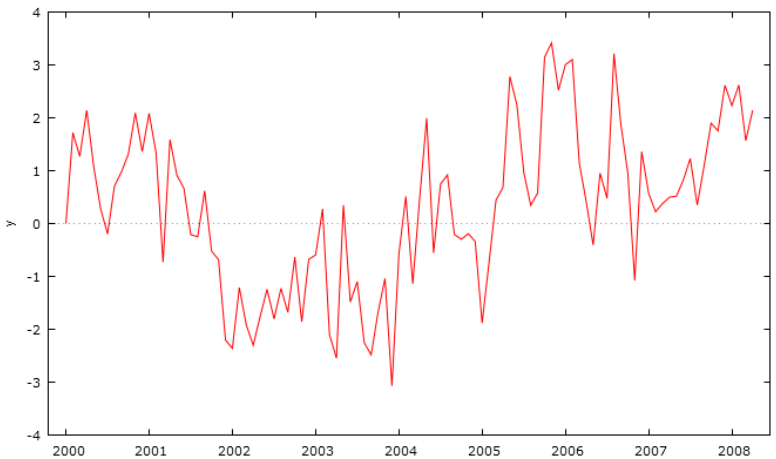
$$\theta(z) = 1 + \theta_1 z + \theta_q z^q \neq 0$$

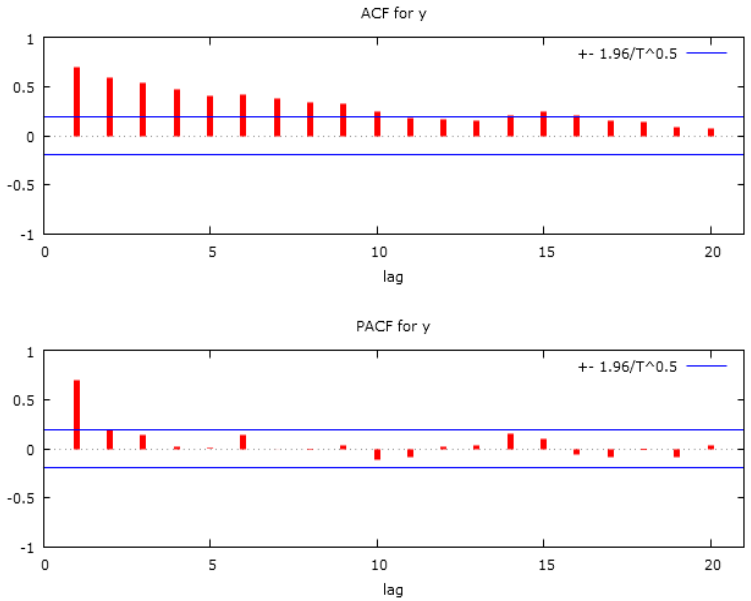
for all $|z| \leq 1$.

Parsimony

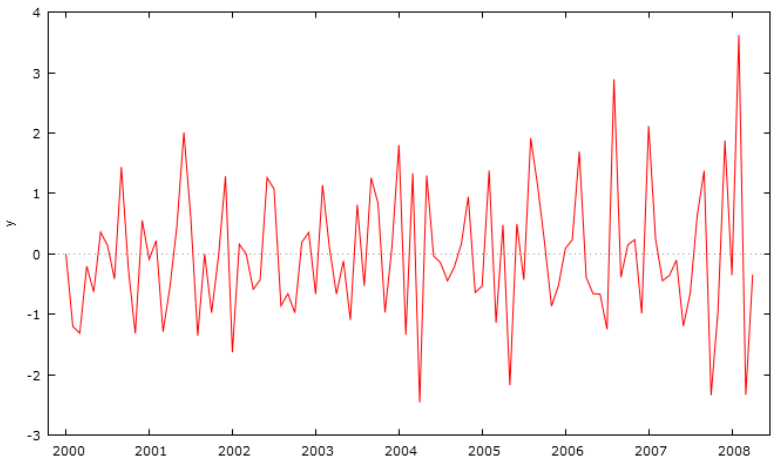
If $\phi(z) = 0$ and $\theta(z) = 0$ share common roots, there exists a stationary and invertible process, $Y_t \sim ARMA(p', q')$, with $p' < p$ and $q' < q$, equivalent to $\{x_t\}$.

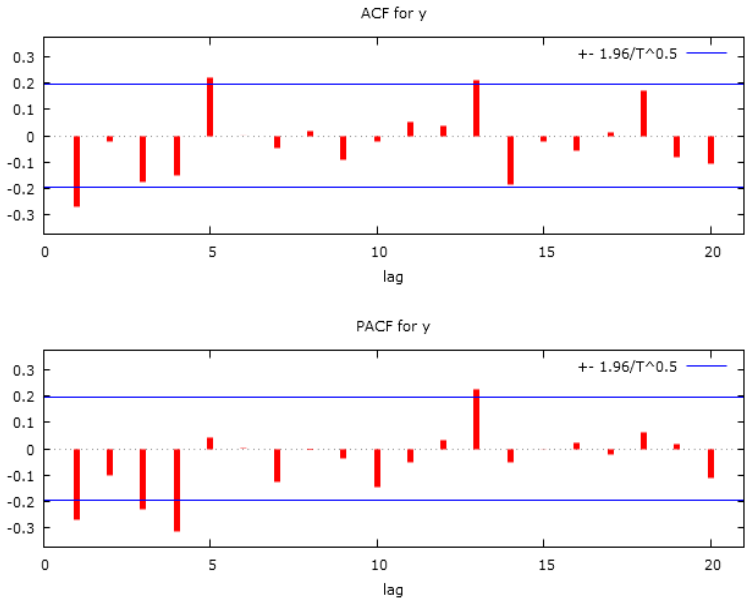
ARMA (1,1) $\phi = 0.9$ and $\theta = -0.5$





ARMA (1,1) $\phi = 0.5$ and $\theta = -0.9$





Estimation of a zero mean AR(1) model

On the assumption that the ϵ_t are independent and come from a normal distribution with mean zero and variance σ^2, the joint probability density function of the $x_1, \ldots, x_n$ is given by

$$L(\phi, \sigma^2) = p(x_1, \ldots, x_n; \phi, \sigma^2) =$$
$$(2\pi\sigma^2)^{-n/2} \times \prod_{t=2}^{n} \exp\left[-\frac{\sum_{t=2}^{n}(x_t - \phi x_{t-1})^2}{2\sigma^2}\right] \times$$
$$(1-\phi^2)^{1/2} \exp\left[\frac{(1-\phi^2)x^2}{2\sigma^2}\right]$$

Maximum likelihood estimates of the parameter (ϕ, σ^2) can be obtained by maximising the likelihood function. Closed-form solutions cannot be found for general ARMA models, but PC does it for us using different algorithms

GOPA GOPA

European Commission

Determining the order of an $ARMA(p, q)$ model

The Box-Jenkins approach to model identification requires a good deal of discretion and experience from the model builder'. However a number of model selection criteria have been proposed

- $AIC(p, q) = -2 \log L(\hat{\phi}, \hat{\theta}, \hat{\sigma}^2) + 2(p + q)n^{-1}$ (Akaike, 1974)
- $BIC(p, q) = -2 \log L(\hat{\phi}, \hat{\theta}, \hat{\sigma}^2) + (p + q)n^{-1} \log n$ (Schwarz, 1978)
- $\Phi(p, q) = -2 \log L(\hat{\phi}, \hat{\theta}, \hat{\sigma}^2) + (p + q)cn^{-1} \log \log n$ (Hannan, 1980)

where $\hat{\sigma}^2$ is the estimate of the error variance σ^2 of the ARMA(p,q) model fitted to the observed time series

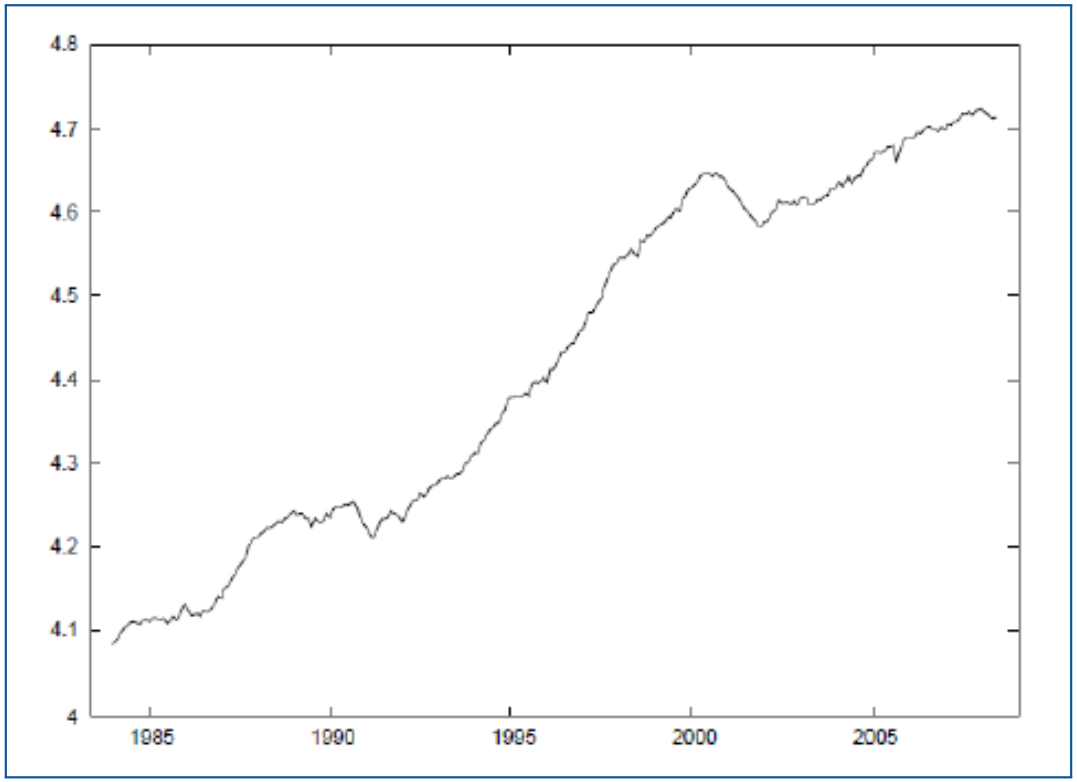
The criteria are used fixing an upper bound for p and q say P and Q and the criteria are elaborated for all $\bar{P} = \{0, 1, \ldots, P)$ and $\bar{Q} = \{0, 1, \ldots, Q)$. Then the 'best' model of order p_1 and q_1 selected by AIC, for example will be

$$AIC(p_1, q_1) = min\ AIC(p, q), p \in \bar{P}\ q \in \bar{Q}$$

GOPA GOPA

European Commission

Illustrative example on USA industrial production



Illustrative example on USA industrial production

Akaike (AIC)

AR \ MA	0	1	2	3	4	5
0	446.06	446.15	439.29	432.49	430.84	430.48
1	445.42	429.84	425.46	427.10	428.69	430.64
2	435.78	426.33	427.27	428.89	430.52	430.42
3	426.36	426.69	428.67	428.62	**421.66**	423.66
4	427.06	428.66	428.71	421.90	423.65	425.60
5	428.84	430.66	426.39	423.69	425.64	426.33

Schwartz (BIC)

AR \ MA	0	1	2	3	4	5
0	453.43	457.20	454.03	450.91	452.94	456.26
1	456.47	444.57	**443.88**	449.20	454.47	460.10
2	450.51	444.75	449.37	454.68	459.99	463.57
3	444.78	448.79	454.45	458.09	454.81	460.49
4	449.16	454.45	458.18	455.06	460.49	466.12
5	454.63	460.13	459.54	460.52	466.15	471.82

GOPA GOPA

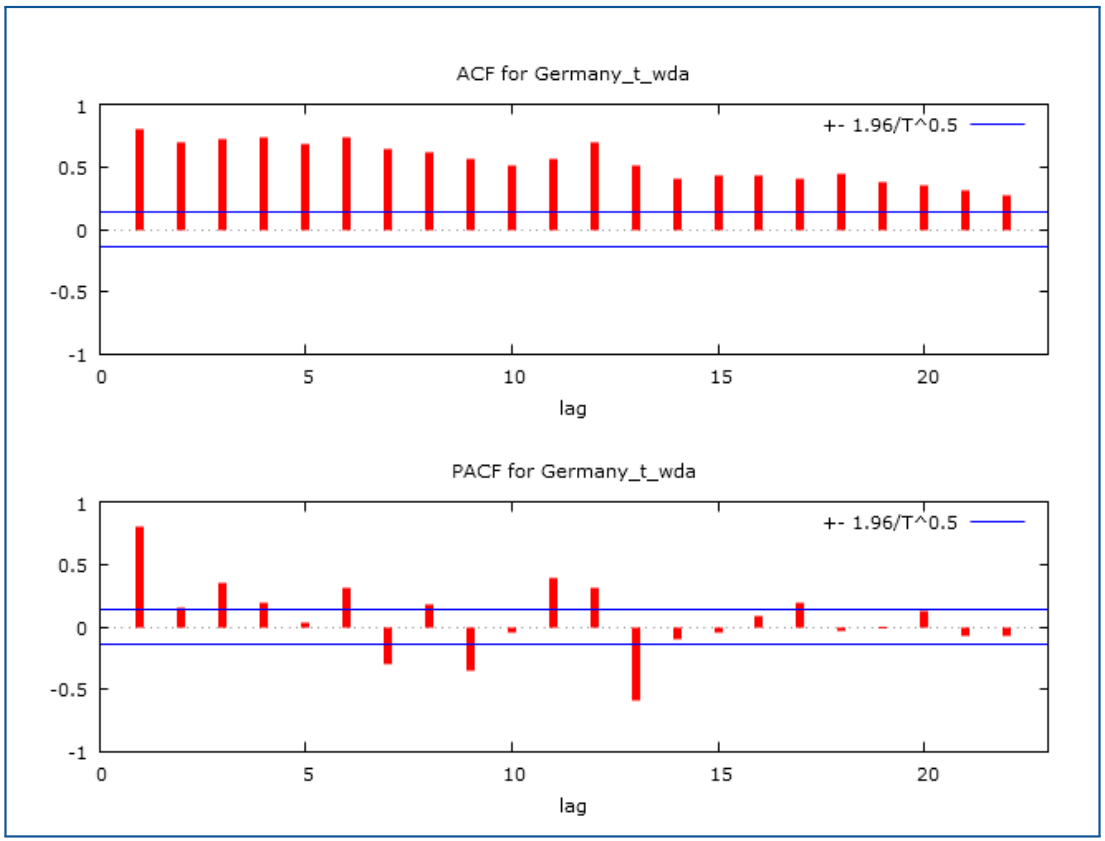
Exercise 1 - 2h

- Introduction to R
- Loading an external dataset
- Simple operation in an interactive mode (lag, difference, log)
- Generate MA, AR, ARMA processes
- first script (program)

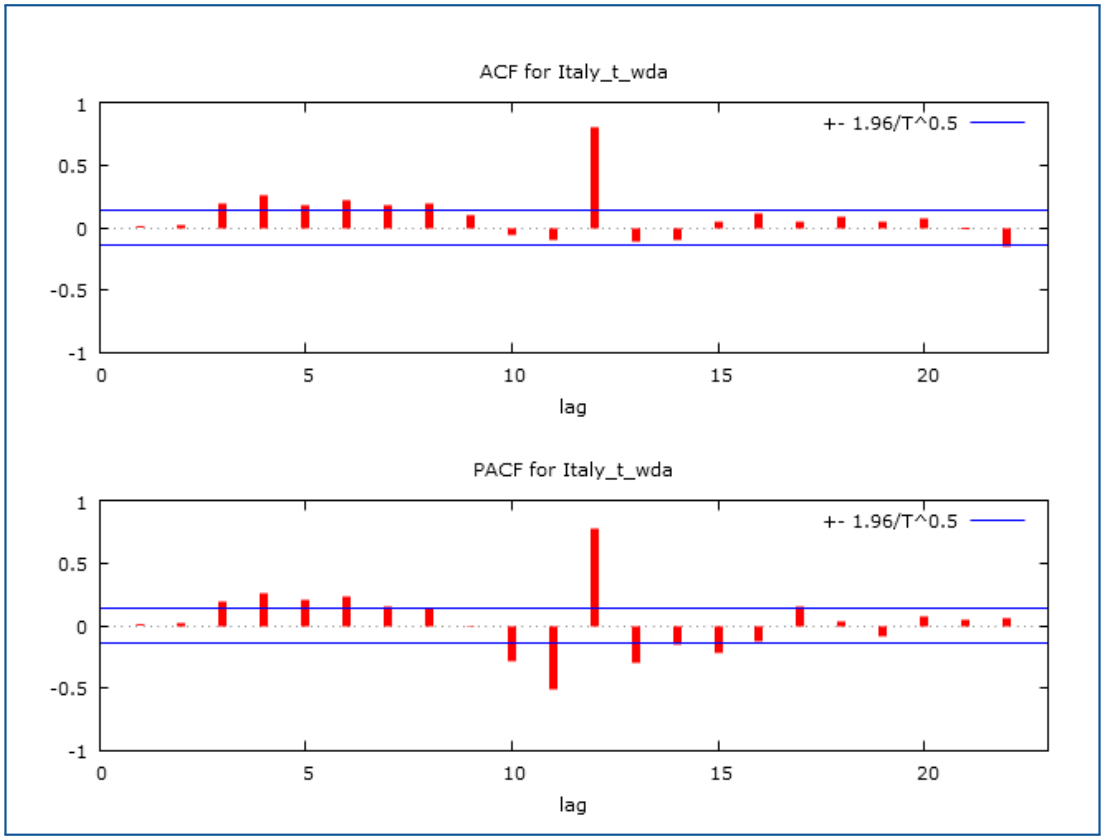
Box-Jenkins's ARIMA loop

- Plot of the series
- Correlogram
- Model identification
- Estimation
- Diagnostics (residuals)

Real data could not follow immediately an ARMA process. Germany monthly index of turnover, WDA



Italy monthly index of turnover, WDA



GOPA GOPA

European Commission

Real time series are different

- Most business and economic time series are far from stationary (eg the economic indicators published by NSO): rarely are mean and/or variance (homoskedastic) stationary
- Non-stationarity emerges from graphic analysis, in particular from timeplot and the correlogram
- In many cases stationarity can be obtained using appropriate transformations
- Before dealing with transformations that induce stationarity is necessary to introduce some notation

Difference operator

- The difference operator Δ is defined as

$$\Delta = (1 - L) \;\rightarrow\; \Delta Y_t = (1 - L) Y_t = Y_t - Y_{t-1} = W_t$$

- W_t is called a differentiated time series
- $\Delta = (1 - L)$ is the difference of order one (difference regular)
- If s represents the seasonality (periodicity) of a time series, then the seasonal difference is:

$$\Delta_s = (1 - L^s)$$

- The regular difference and the seasonal difference of d and D respectively, are indicated with

$\Delta^d = (1 - L)^d$
$\Delta_s^D = (1 - L^s)^D$

Time series non-stationary in mean

- A time series can be made mean stationary by applying the difference operator

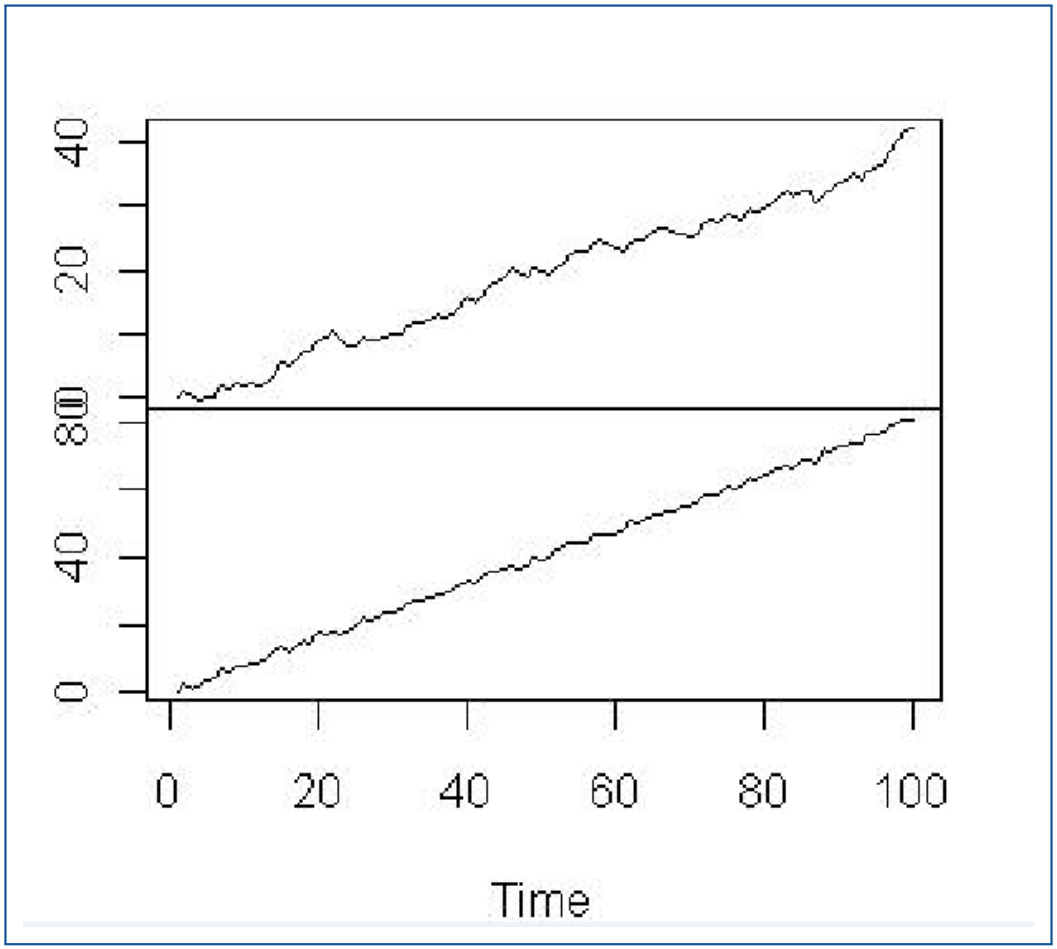
$$\Delta^d Y_t = (1 - L)^d Y_t = W_t \quad \rightarrow \quad \text{stationary}$$

- In general $d = 1$, sometimes $d = 2$

GOPA GOPA

European Commission

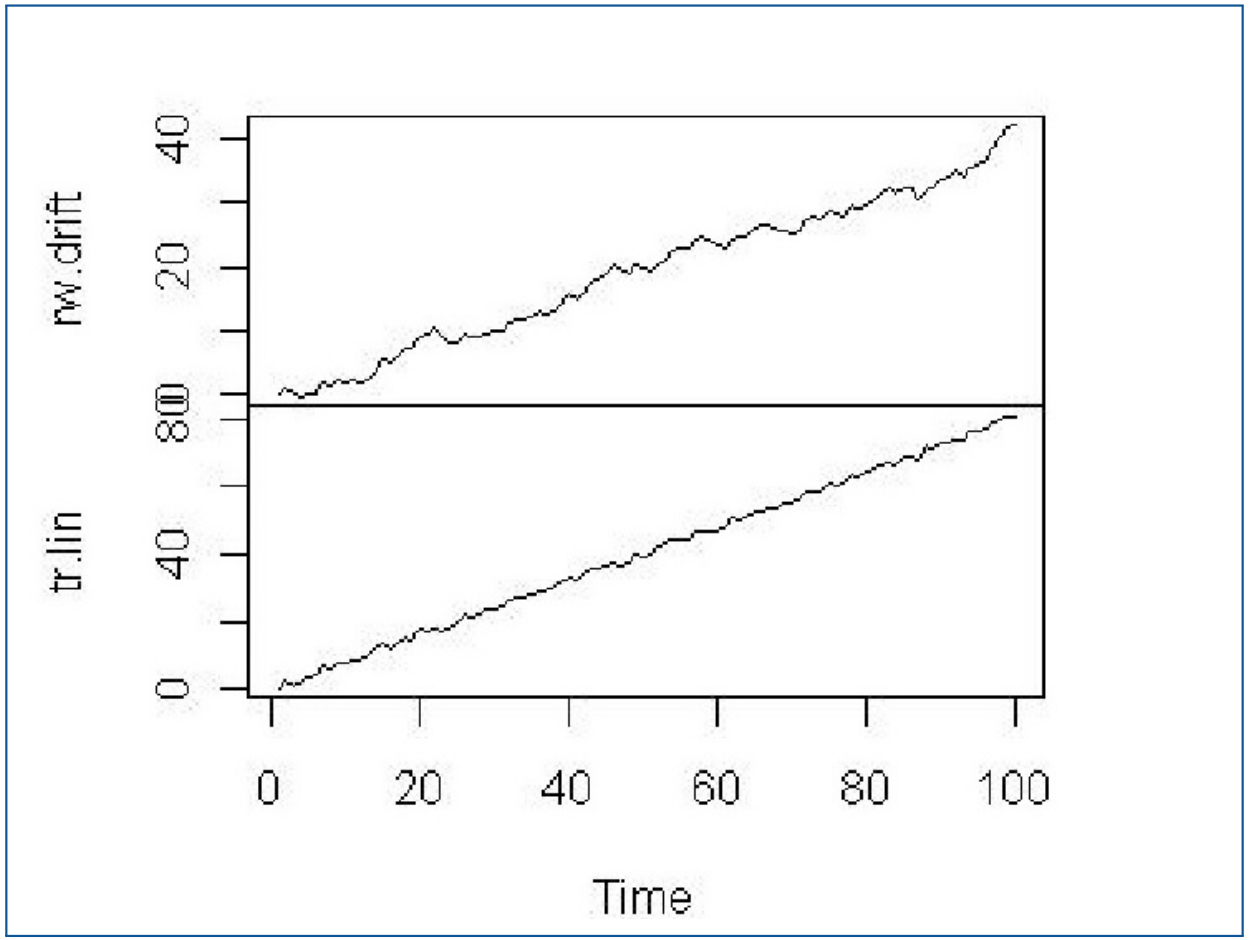
Are time series similar?



GOPA GOPA

European Commission

The two time series are not similar



I as Integrated

$$ARMA(p,q) : \Phi(B)x_t = \Theta(B)a_t$$

$$ARIMA(p,d,q) : \Phi(B)\Delta^d x_t = \Theta(B)a_t$$

d=0 o d=1

Results

- Random walk, integrated time series of order 1 (Difference stationarity) are synonyms;
- For integrated time series of order 1 a linear trend should not be estimated;
- **ACF of an integrated series declines very slowly**.

GOPA GOPA

European Commission

Deterministic Trend vs stochastic trend

Deterministic trends need to be incorporated in the statistical analysis for example by introducing a linear trend in the model.

However, attention should be paid to the presence of non-stationarty induced by persistent accumulation of past disturbances, defined as unit root processes.
Such processes can be fitted to time series characterised by stochastic trends, i.e. trends evolving randomly in time.

A simple model

$$y_t = y_0 + \beta t + u_t$$

- Simple regression model for a variable y_t containing a linear deterministic trend with a slope given by β and generated by an initial value y_0;
- The u_t error is an autoregressive model of order 1 with random data shocks given by ϵ_t with distribution $WN[0, \sigma^2]$.

The model in its economic interpretation supposes the time series follows a steady growth trend over time (linear function of the time) to which an economic cycle effect is added *short term effect* that is represented by a stationary process (if $|\rho| < 1$).

Deterministic trend

$$u_t = \epsilon_t + \rho\epsilon_{t-1} + \rho^2\epsilon_{t-2} + \dots \tag{1}$$

$$u_t = \rho u_{t-1} + \epsilon_t \quad \left(u_t = \frac{1}{1-\rho L}\ \epsilon_t\right) \tag{2}$$

u_t errors are the sum of all previous shocks ϵ_{t-i} but the effect of previous disturbances decreases with time because $|\rho| < 1$. However, the process y_t is not a stationary process because $E(y_t) = y_0 + \beta t$ is not constant over time (for $\beta \neq 0$).
However if we consider trend variations around the trend the process u_t is stationary and, therefore, the y_t is defined as a stationary process around a trend **TS processes, Trend stationary**.

Stochastic trend

If $\rho = 1$ then $u_t = \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2} + \ldots$, so every disturbance persists indefinitely and has a permanent effect on u_t. Consequently it is said that u_t has the stochastic trend $\sum_{i=1}^{t} \epsilon_i$.

The difference between a stochastic trend and a linear deterministic trend is that the increases of a stochastic trend are random, whereas those of a deterministic trend are constant over time (since the sum in continuous time corresponds to the concept of integral, the process itself is defined **integrated of order 1**).

$$
\begin{aligned}
Y_t &= 0,5Y_{t-1} + \epsilon_t => (1 - 0,5L)Y_t = \epsilon_t \\
Y_t &= Y_{t-1} + \epsilon_t => (1 - L)Y_t = \Delta Y_t = \epsilon_t \\
Y_t &= Y_{t-1} + \mu + \epsilon_t => (1 - L)Y_t = \mu + \epsilon_t
\end{aligned}
$$

Definition:

Y_t is a random walk if ΔY_t is a white noise. It can easily verified that:

$$
\begin{aligned}
Y_1 &= Y_0 + \epsilon_1 \\
Y_2 &= Y_1 + \epsilon_2 = Y_0 + \epsilon_1 + \epsilon_2 \ldots \\
Y_t &= \sum_{i=1}^{t} \epsilon_i
\end{aligned}
$$

Implication

- mean of the process is again equal to zero
- variance of Y_t is $t\sigma^2$, so it increases along the time and stationarity does not hold

Characteristic

- there is a relevant increase of the persistence of the process
- this implies stochastic trends

GOPA GOPA

European Commission

ARIMA

$$(1-L)Y_t = W_t \sim stationary$$

$$(1-\phi L)W_t = (1-\theta L)\epsilon_t \sim ARMA(1,1)$$

$$(1-\phi L)(1-L)Y_t = (1-\theta L)\epsilon_t \sim ARIMA(1,1,1)$$

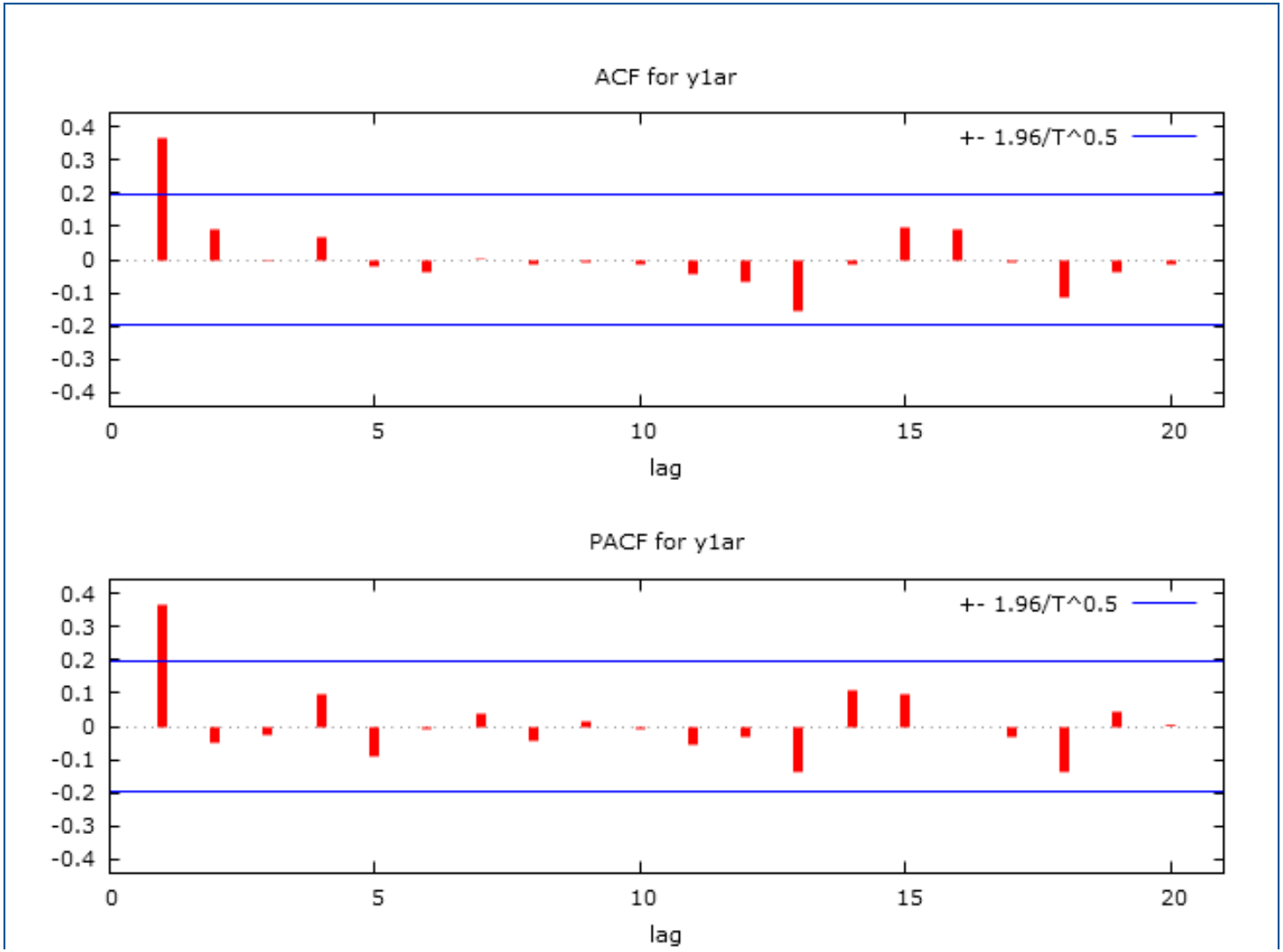
more general we refer to ARIMA(p,d,q)

how to check for integration

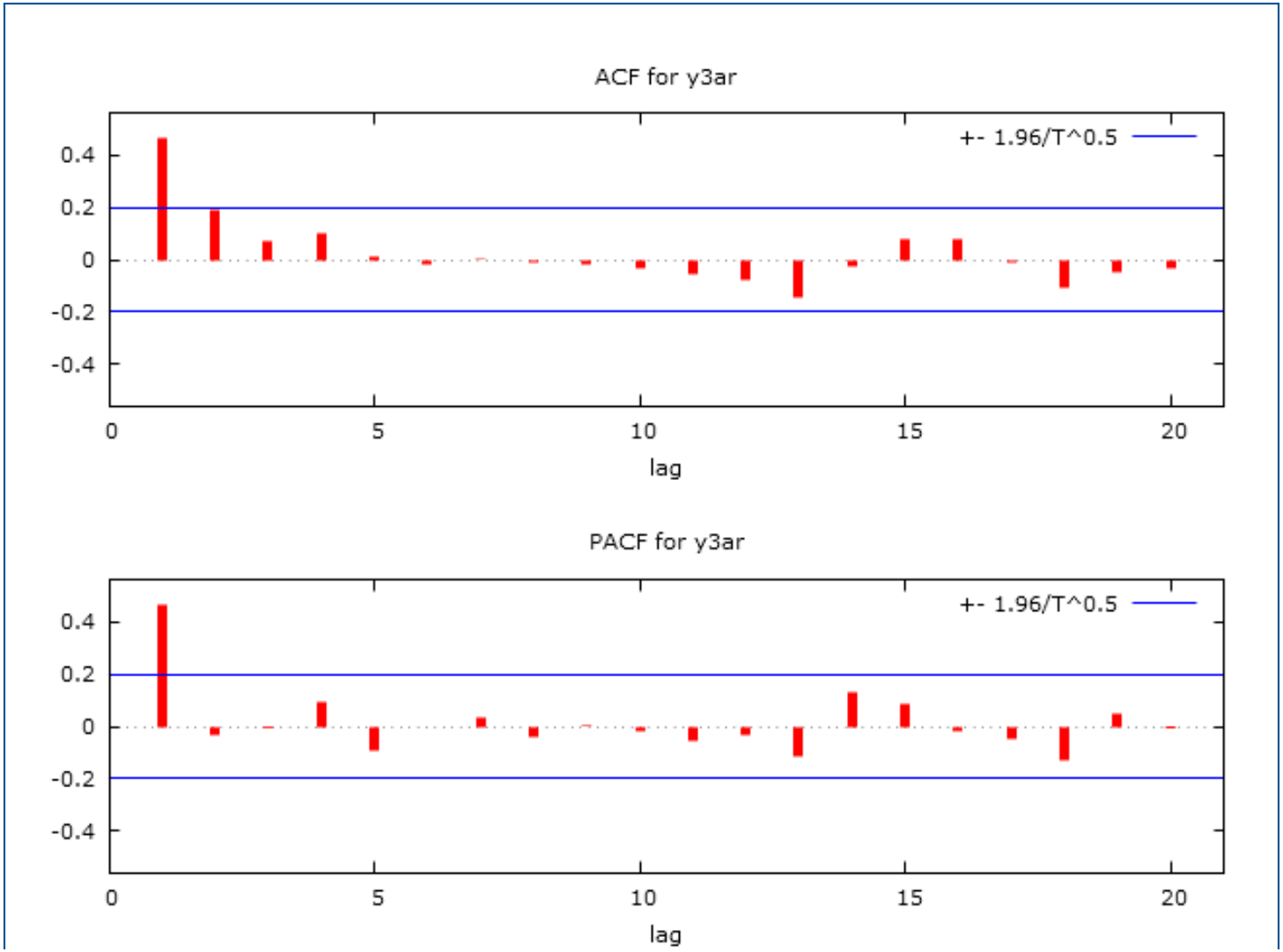
two ways

- graphic analysis (heuristical approach BJ)
- test DF or ADF

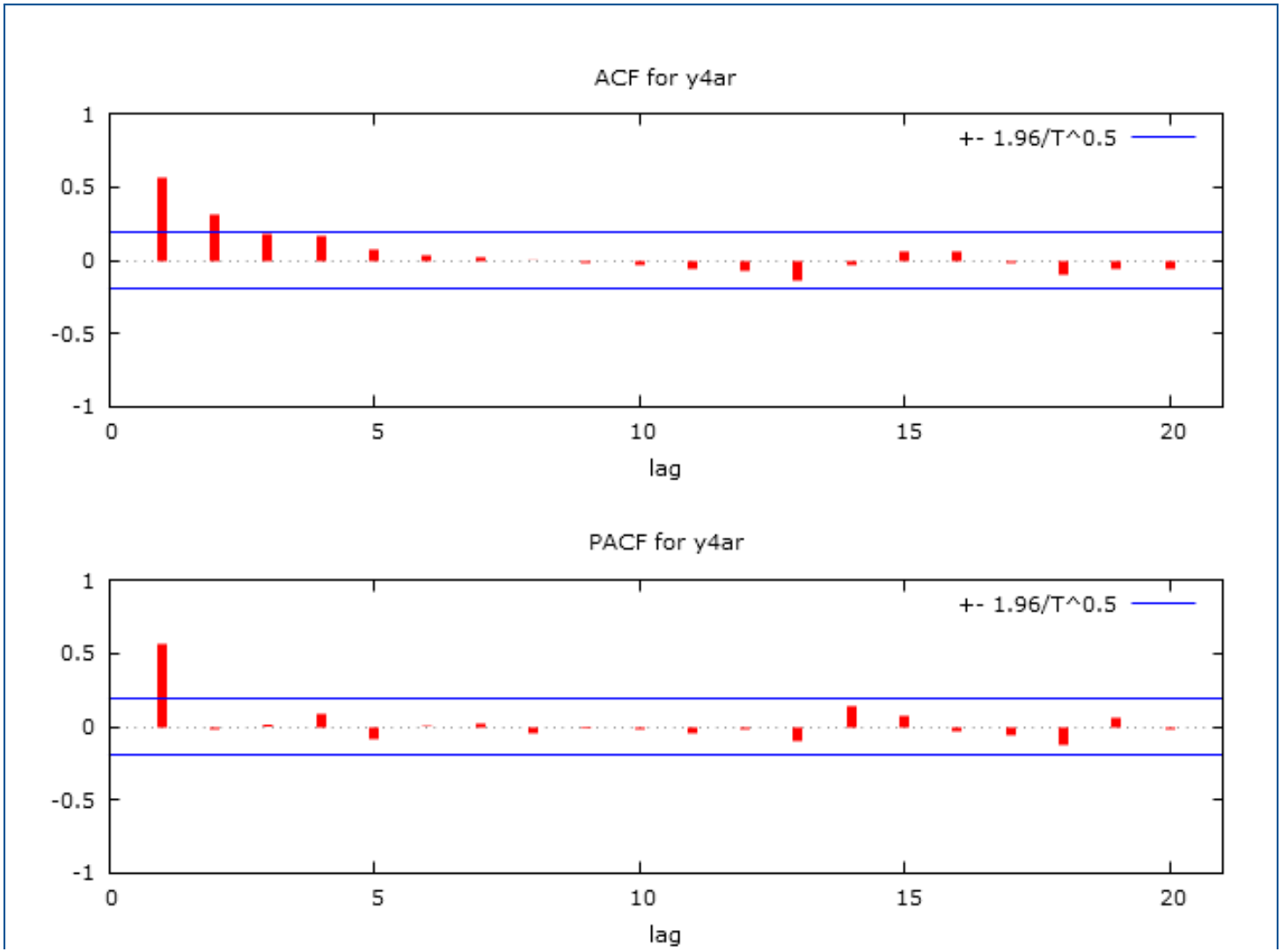
graphich characteristic of AR process AR(1) $\phi = 0.5$



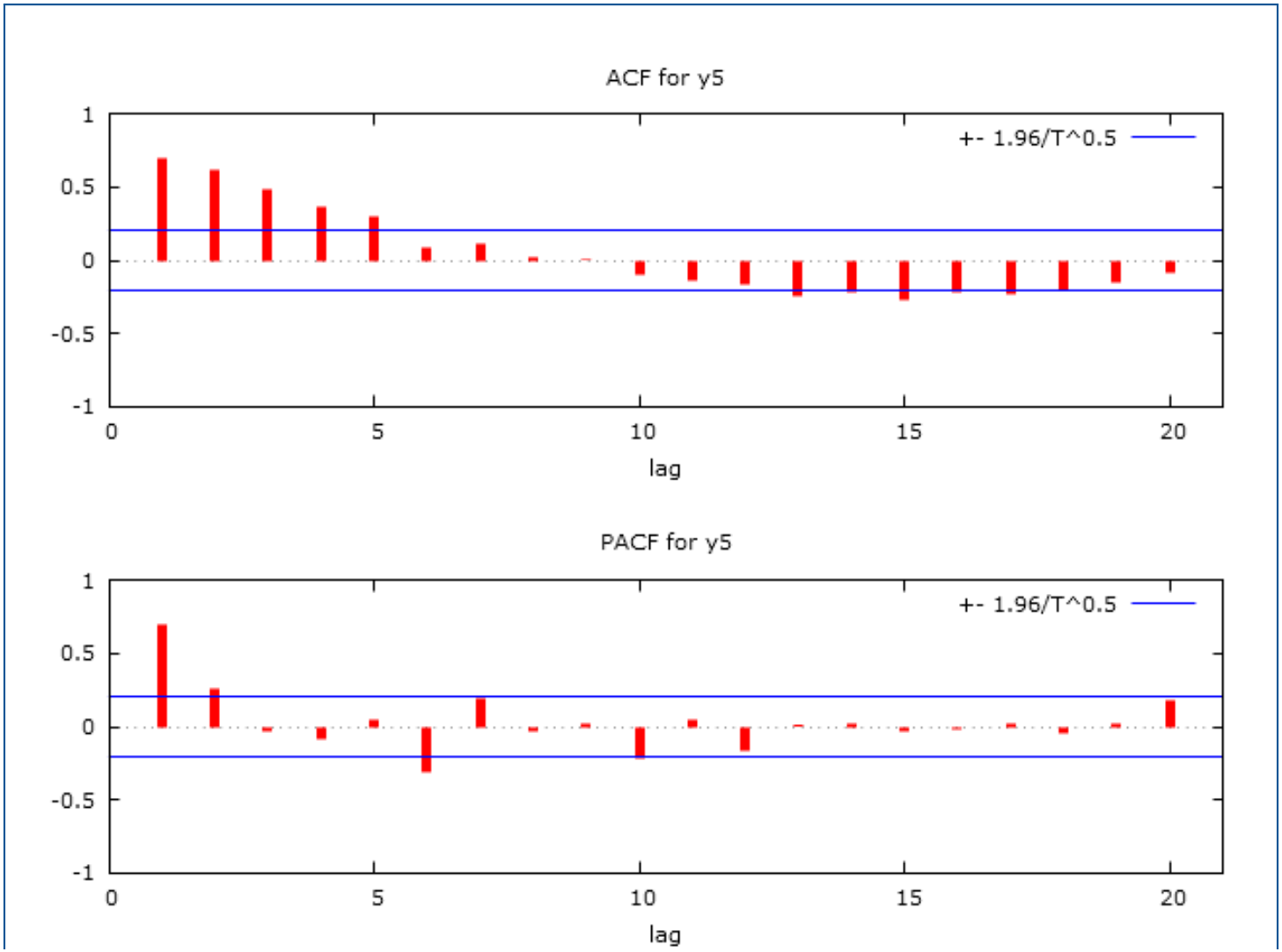
AR(1) $\phi = 0,6$


ACF for y3ar
+- 1.96/T^0.5
0.4
0.2
0
-0.2
-0.4
0
5
10
15
20
lag
PACF for y3ar
+- 1.96/T^0.5
0.4
0.2
0
-0.2
-0.4
0
5
10
15
20
lag

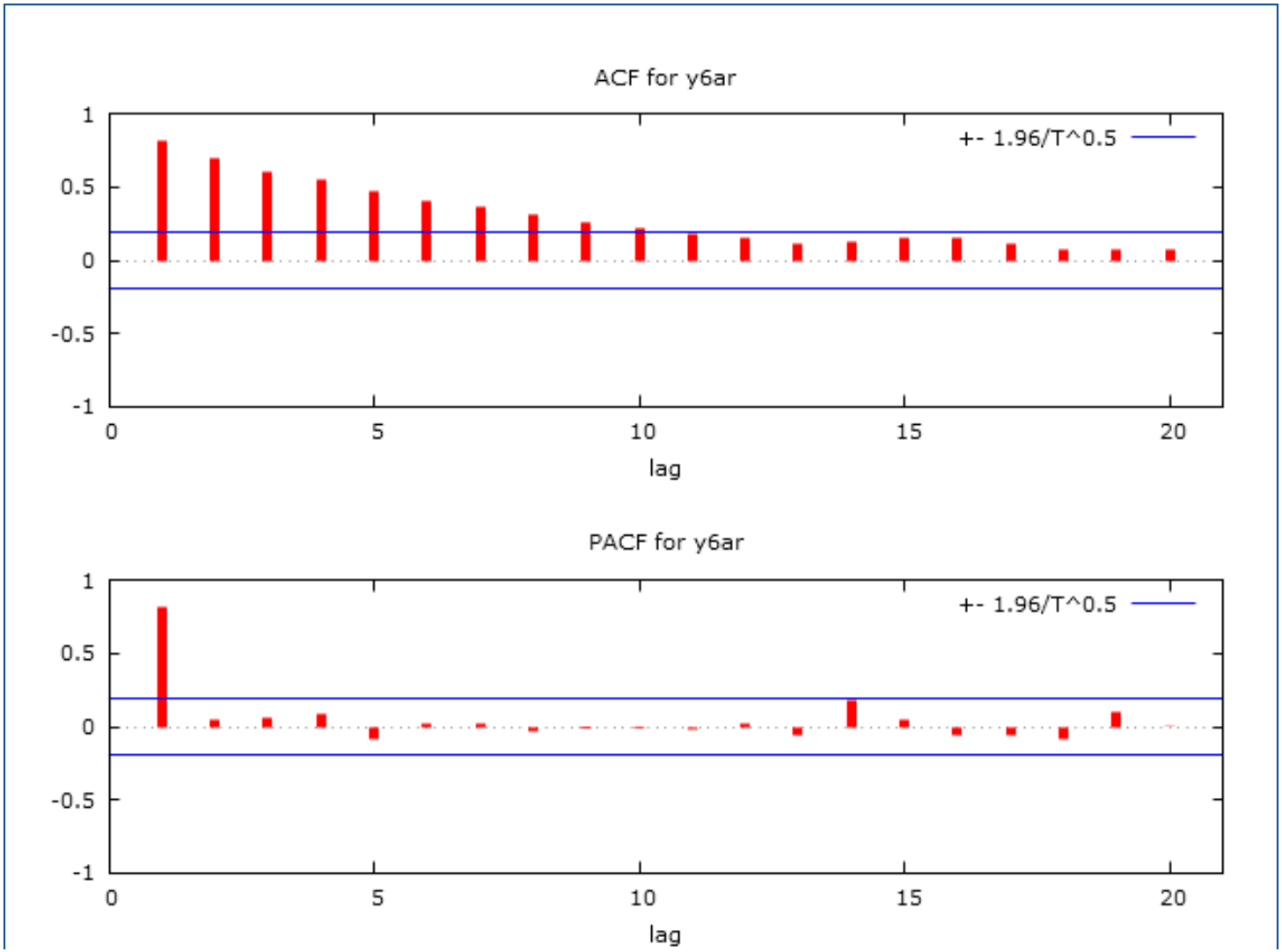
AR(1) $\phi = 0,7$


ACF for y4ar
+- 1.96/T^0.5
1
0.5
0
-0.5
-1
0
5
10
15
20
lag
PACF for y4ar
+- 1.96/T^0.5
1
0.5
0
-0.5
-1
0
5
10
15
20
lag

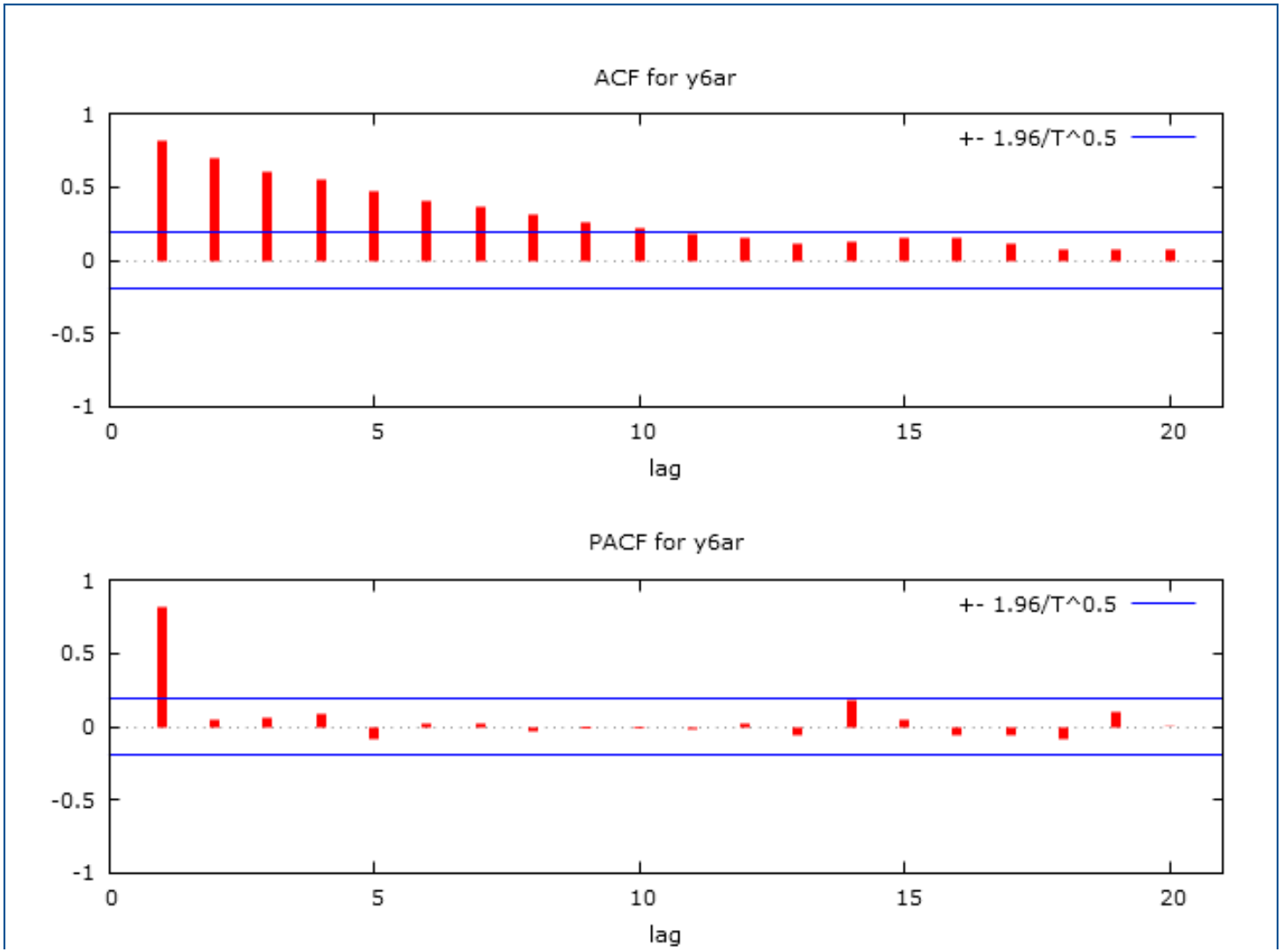
AR(1) $\phi = 0,8$


ACF for y5
+- 1.96/T^0.5
1
0.5
0
-0.5
-1
0
5
10
15
20
lag
PACF for y5
+- 1.96/T^0.5
1
0.5
0
-0.5
-1
0
5
10
15
20
lag

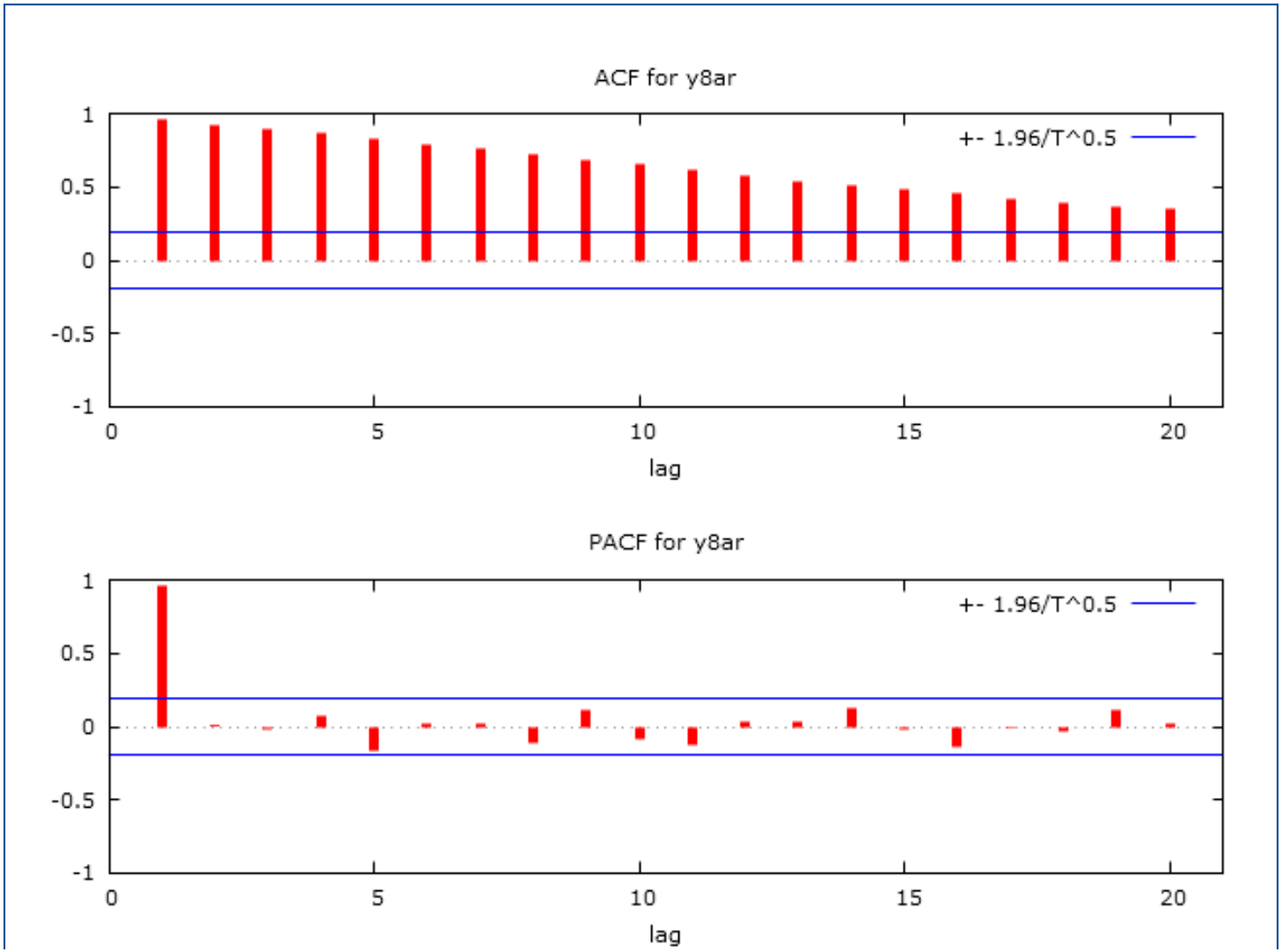
AR(1) $\phi = 0,9$


ACF for y6ar
+- 1.96/T^0.5
1
0.5
0
-0.5
-1
0
5
10
15
20
lag
PACF for y6ar
+- 1.96/T^0.5
1
0.5
0
-0.5
-1
0
5
10
15
20
lag

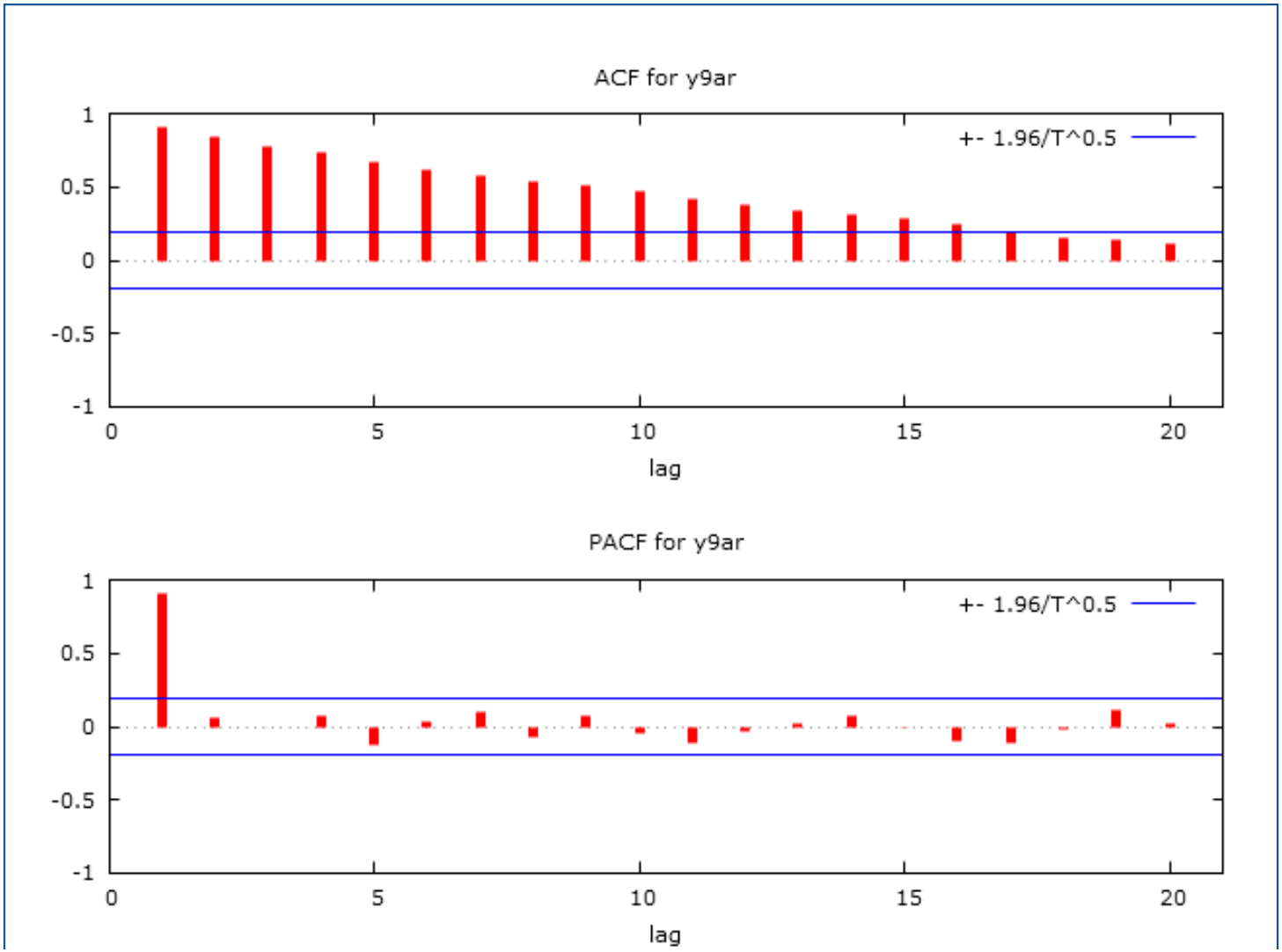
AR(1) $\phi = 0,99$


ACF for y6ar
+- 1.96/T^0.5
1
0.5
0
-0.5
-1
0
5
10
15
20
lag
PACF for y6ar
+- 1.96/T^0.5
1
0.5
0
-0.5
-1
0
5
10
15
20
lag

AR(1) $\phi = 1$


ACF for y8ar
+- 1.96/T^0.5
1
0.5
0
-0.5
-1
0
5
10
15
20
lag
PACF for y8ar
+- 1.96/T^0.5
1
0.5
0
-0.5
-1
0
5
10
15
20
lag

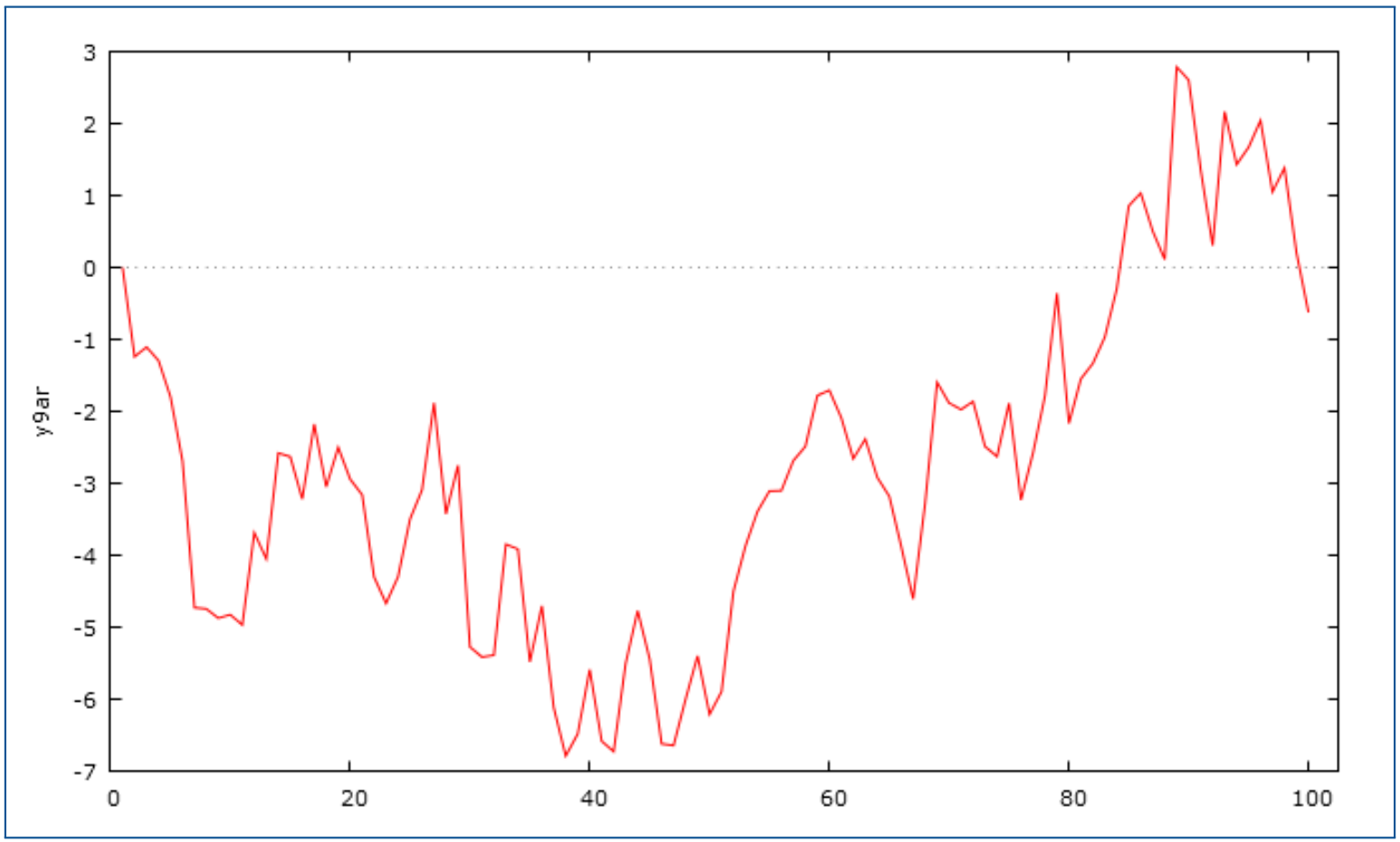
AR(1) $\phi = 1,02$


ACF for y9ar
+- 1.96/T^0.5
1
0.5
0
-0.5
-1
0
5
10
15
20
lag
PACF for y9ar
+- 1.96/T^0.5
1
0.5
0
-0.5
-1
0
5
10
15
20
lag

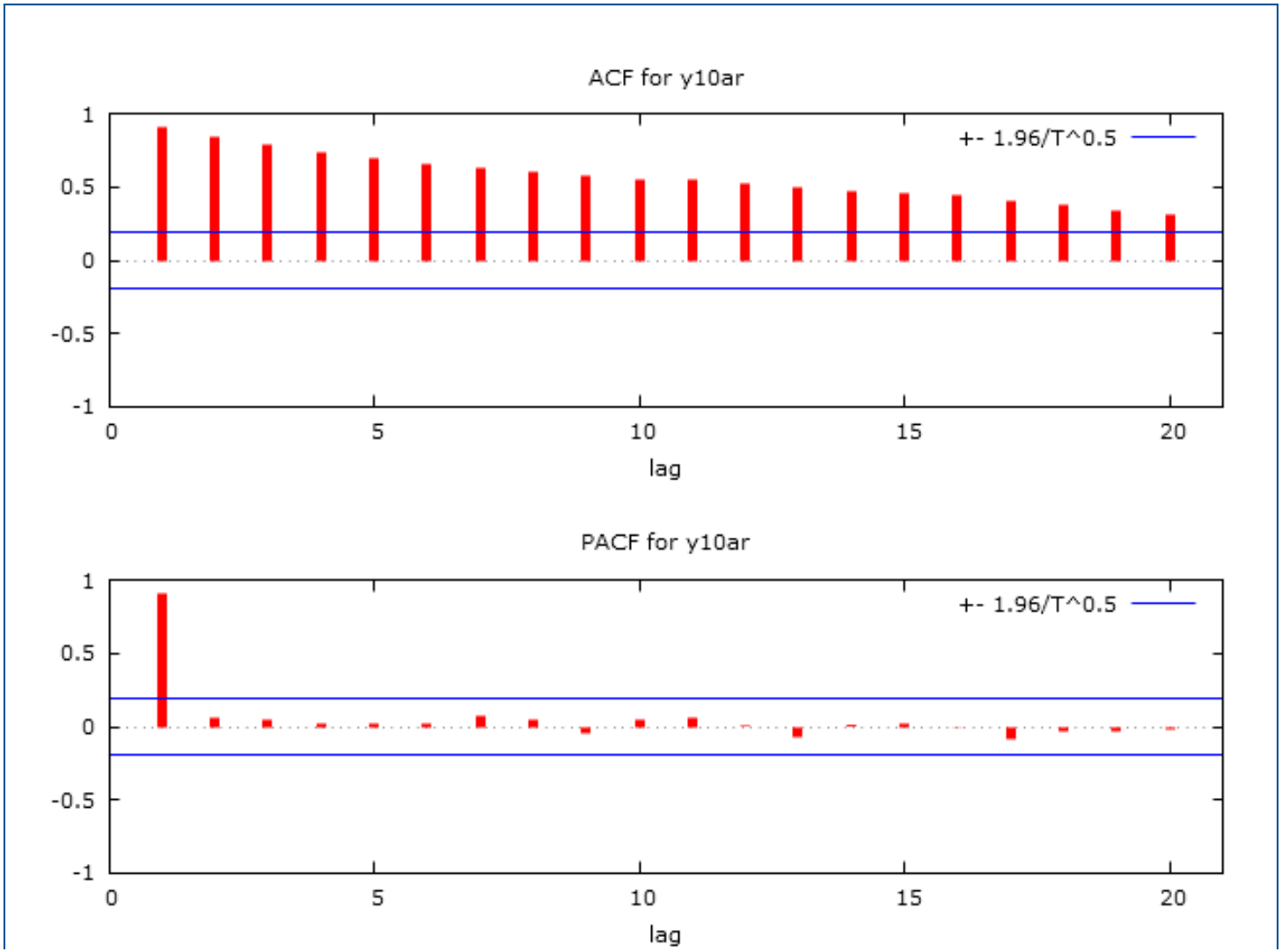
GOPA GOPA

European Commission

AR(1) $\phi = 1,02$



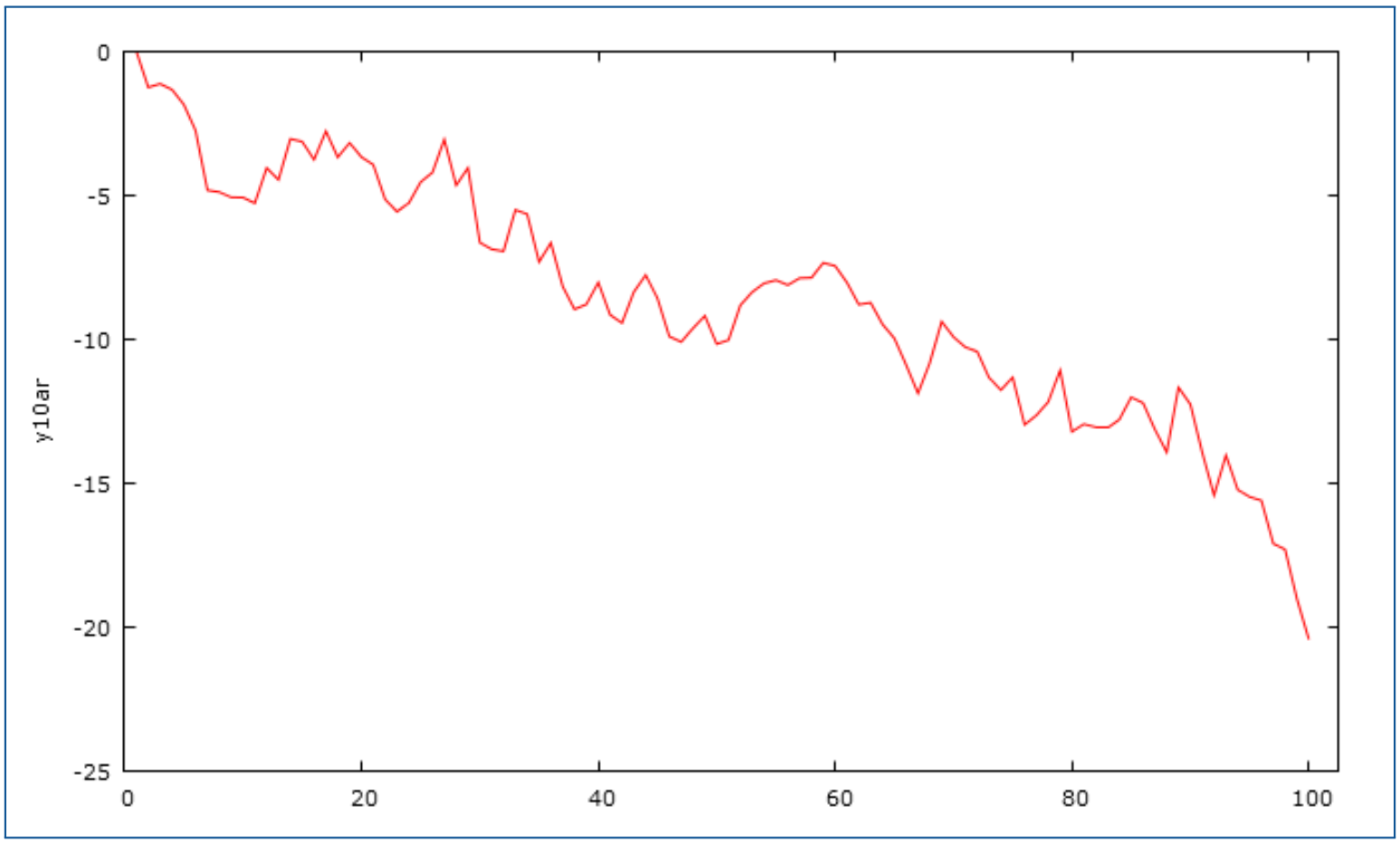
AR(1) $\phi = 1,03$


ACF for y10ar
+- 1.96/T^0.5
1
0.5
0
-0.5
-1
0
5
10
15
20
lag
PACF for y10ar
+- 1.96/T^0.5
1
0.5
0
-0.5
-1
0
5
10
15
20
lag

GOPA GOPA

European Commission

AR(1) $\phi = 1,03$



Using arima.sim:

1. Simulate:

$$x_t = 0.999999 x_{t-1} + \epsilon_t$$

2. Simulate:

$$y_t = y_{t-1} + \epsilon_t - 0.999999 \epsilon_{t-1}$$

It is everything straightforward?

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

Subtracting y_{t-1} we get:

$$\Delta y_t = \rho y_{t-1} + \epsilon_t$$

with ϵ_t white noise and $\rho = \phi - 1$. So since the previous relationship is similar to a regression model the null hypothesis on parameter $\rho = 0$ can be tested by the statistic:

$$t_\rho = \frac{\hat{\rho}}{\sqrt{\hat{Var}(\hat{\rho})}}$$

Is equivalent to test the null hypothesis of non-stationarity.

1. The asymptotic distribution of t_ρ is not standard (Series is not stationary) but Dickey and Fuller have calculated the critical values;
2. It is possible to consider an augmented Dickey-Fuller test keeping into account the short-term persistence of ∇y_t using an AR (p) model the null hypothesis of non-stationary is equivalent to checking for model:

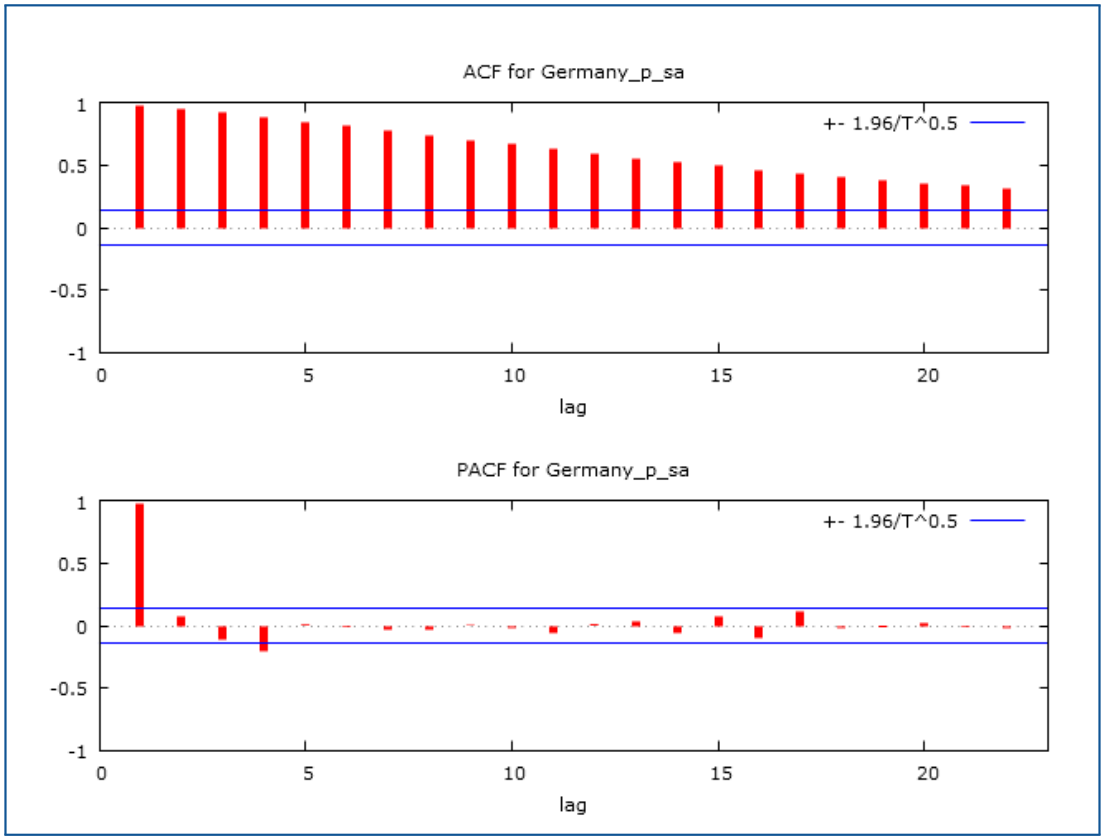
$$\Delta y_t = \rho y_{t-1} + \gamma_1 \Delta y_{t-1} + \ldots + \gamma_p \Delta y_{t-p+1} + \epsilon_t$$

   the hypothesis $\rho = \phi - 1 = 0$
3. In some econometric packages the test it is assumed that in the previous parametrization there is drift and/or a deterministic trend
4. In `urca` package *ur.df*.

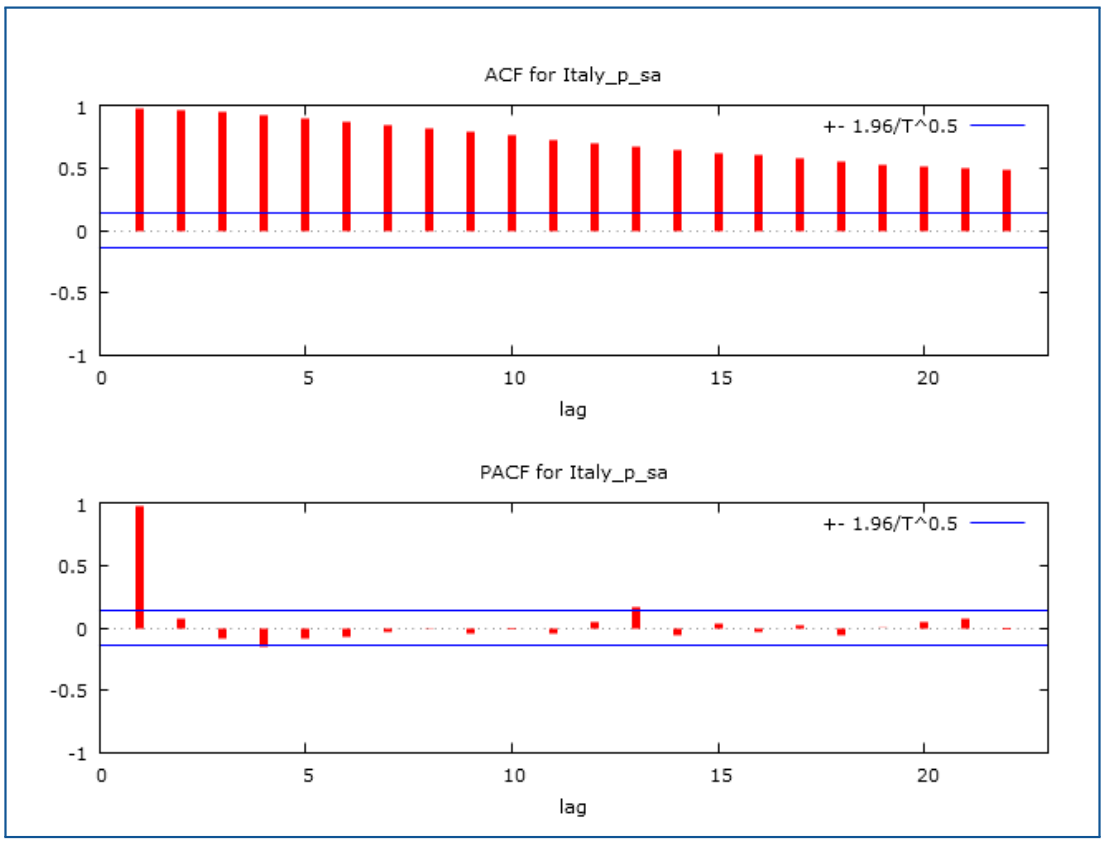
Exercise 2 - 1h

- unit root test
- exercise on interactive mode
- write a script for the test

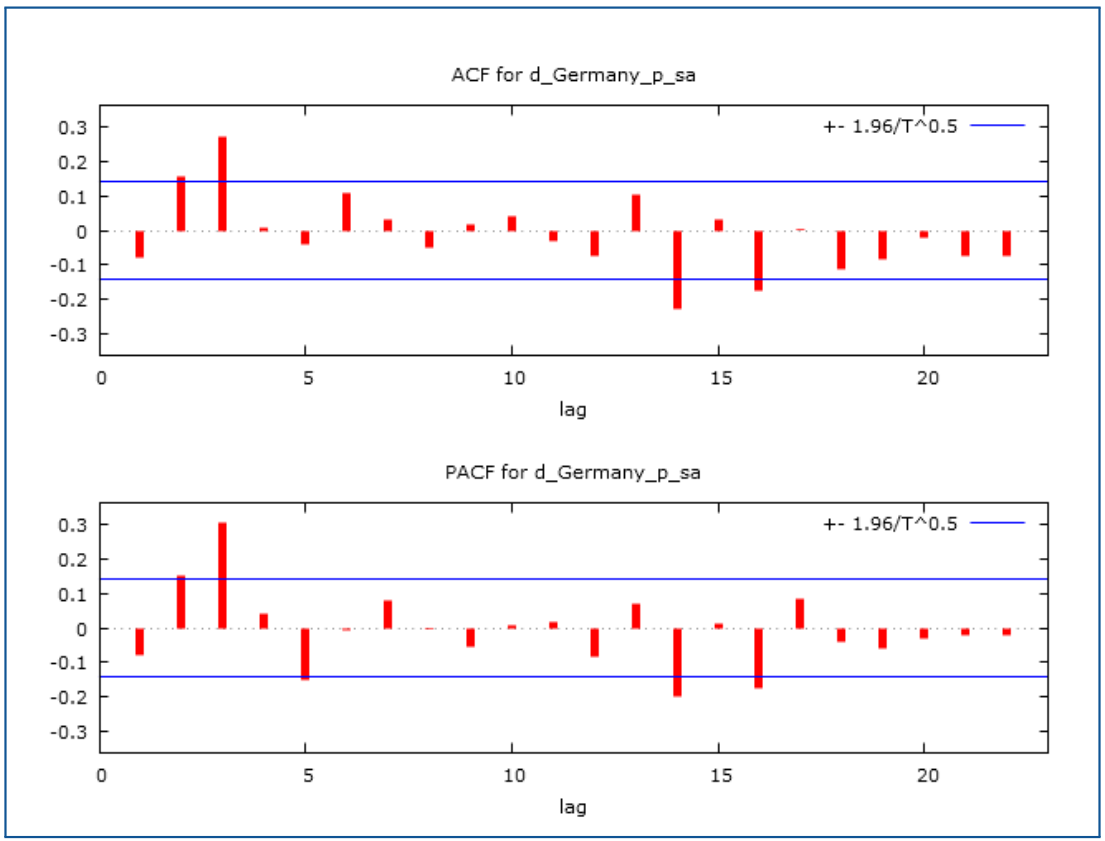
again real data could not follow immediately an ARIMA process.
Monthly index of production in industry (Germany), SA



Monthly index of production in industry (Italy), SA



Monthly index of production in industry (Italy), SA, first difference



Difference between seasonally adjusted and seasonally unadjusted data

Example

Suppose to identify a model for italian production in August or sales in December. Both evidences are clearly only if compared with the same month in previous year

Period: $s, 2s, \ldots Ps$. AR and MA refer to the seasonal frequency

$$\Phi(L^s) = 1 - \Phi_1 L^s - \Phi_2 L^{2s} \ldots - \Phi_P L^{Ps}$$

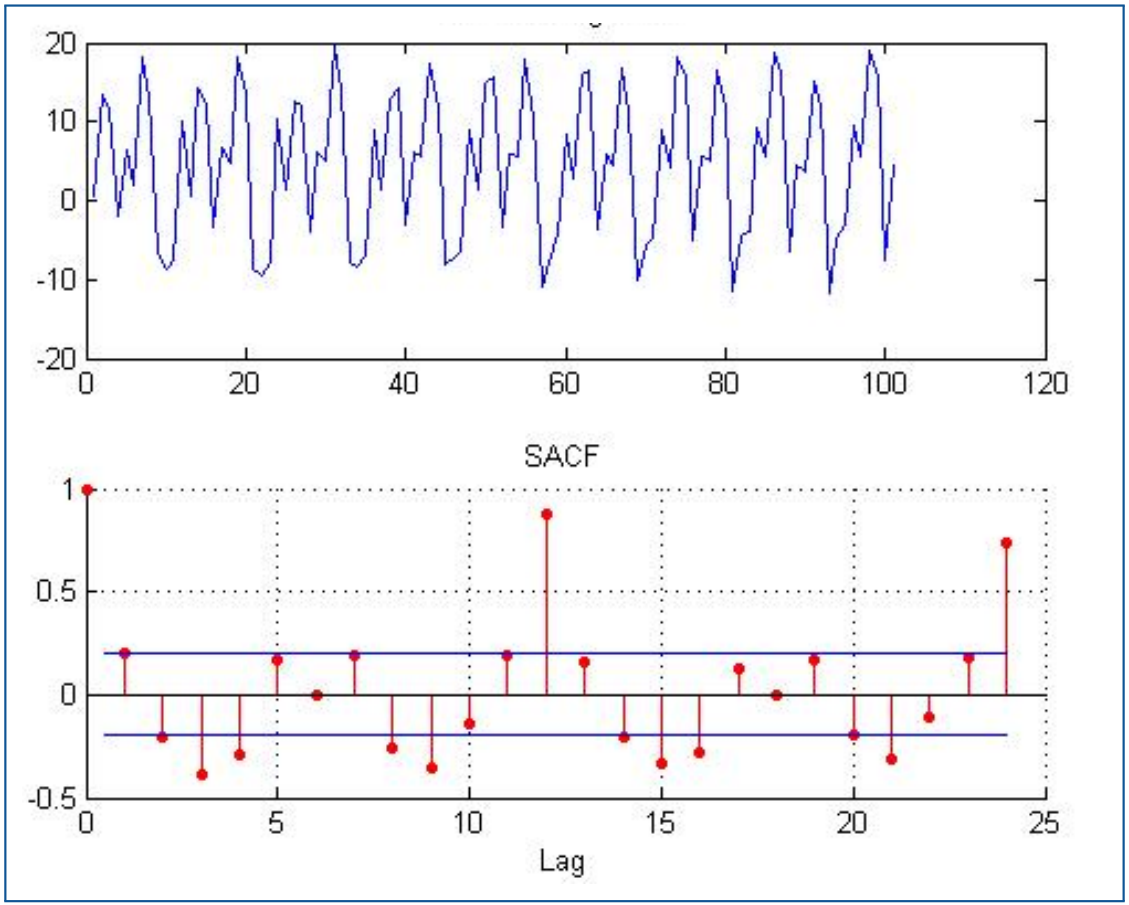
$$\Theta(L^s) = 1 - \Theta_1 L^S - \Theta_2 L^{2s} \ldots - \Theta_Q L^{Qs}$$

$$(1 - L^{12})Y_t = \epsilon_t \ (model1)$$
$$(1 - L)(1 - L^{12})Y_t = \epsilon \ (model2)$$

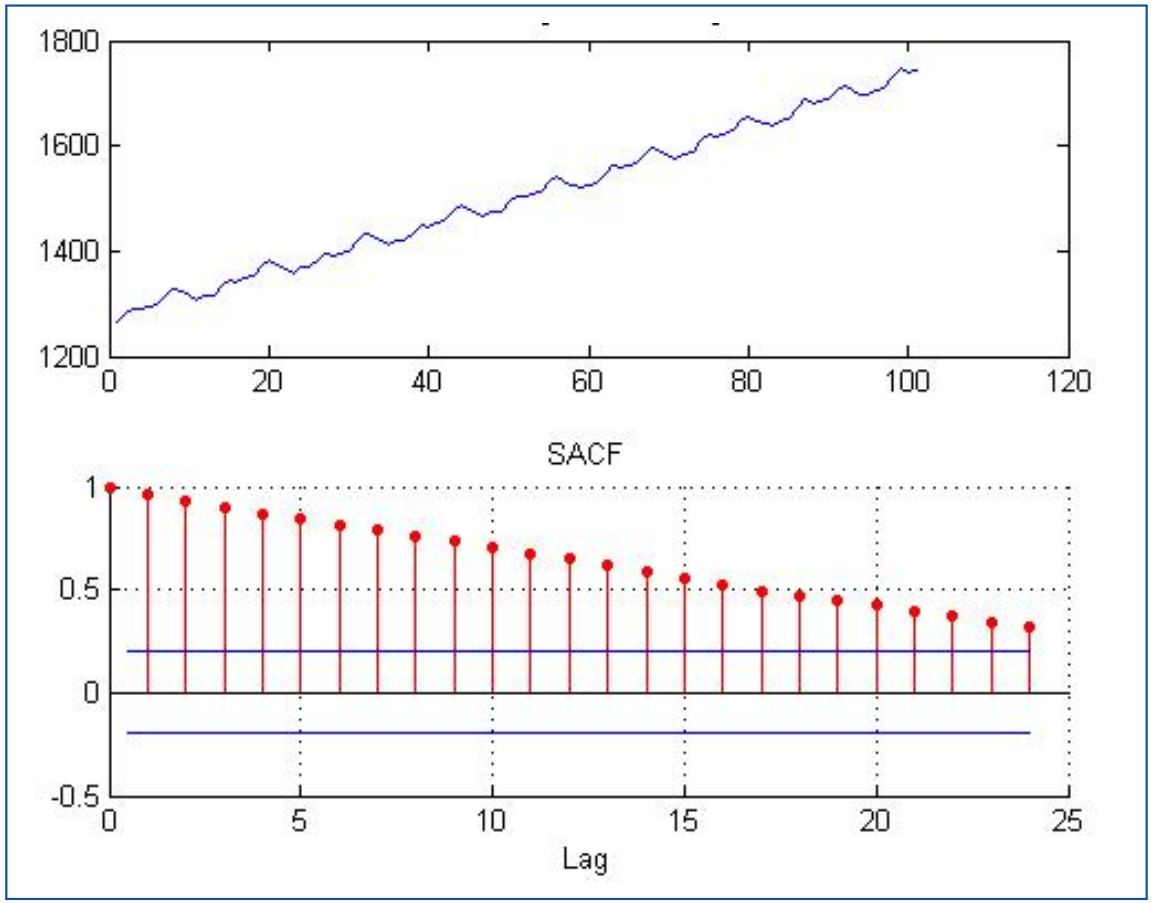
GOPA GOPA

European Commission

Simulation: model 1



Simulation: model 2



How to manage seasonal integration

two ways

- graphical analysis (informal approach BJ)
- test extension ADF (HEGY, 1990)

Notation SARIMA: example

Multiplicative ARMA are popular because they identify the single components

$$\phi(L)\Phi(L)\Delta\Delta_s Y_t = \theta(L)\Theta(L)\epsilon_t$$

$ARIMA(p,1,q)(P,1,Q)_s$

We suppose that the component a_t evolves with a seasonal autocorrelation

$$\phi(L)\Delta Y_t = \theta(L)\epsilon_t$$

with ϵ_t such that

$$\Phi(L)\epsilon_t = \Theta(L)e_t$$

and then

$$\phi_p(L)\Phi_P(L)\Delta(\Delta_s)Y_t = \theta_q\Theta_Q(L)e_t$$

Basic concepts

- Defining the information set $\Im_t$ 'consisting of data together with knowledge, theories, or assumptions about the properties about the process as available at time T. For example $\Im_t$ consists of a sample of previous values of the series together with the information that it has zero mean and the assumption that the series is stationaty
- stochastic characteristics of the process
- define a cost function
- point interval prediction

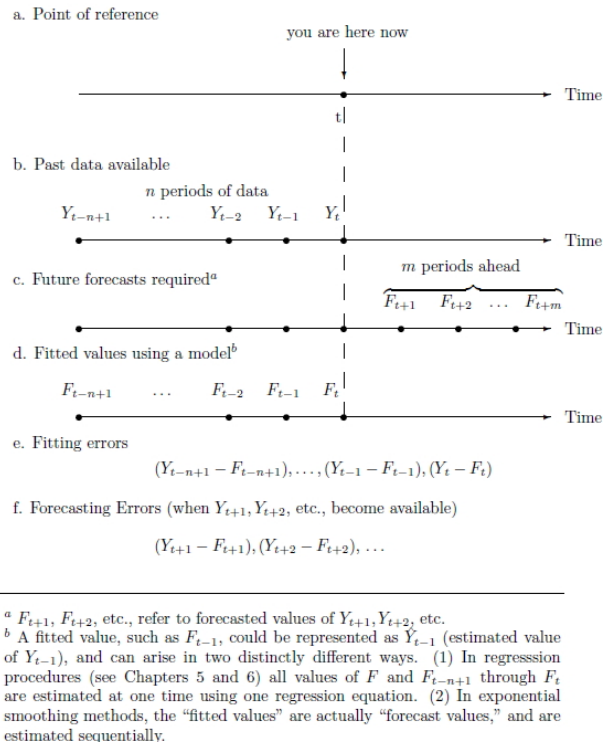
Moving average and forecasting

Generalizing two categories of forecasting methods:

- Averaging Methods
  - Simple arithmetic mean
  - Moving averages
- Exponential smoothing methods

Let us see the differences

Forecasting scenario



[a] F_{t+1}, F_{t+2}, etc., refer to forecasted values of Y_{t+1}, Y_{t+2}, etc.
[b] A fitted value, such as F_{t-1}, could be represented as $\hat{Y}_{t-1}$ (estimated value of Y_{t-1}), and can arise in two distinctly different ways. (1) In regresssion procedures (see Chapters 5 and 6) all values of F and F_{t-n+1} through F_t are estimated at one time using one regression equation. (2) In exponential smoothing methods, the "fitted values" are actually "forecast values," and are estimated sequentially.

Source: Figura 4.2 in Makridakis, S. Wheelwright, S. Hyndman R. "Forecasting: Methods and Applications 3rd edition" pag.139

Averages

- Simple arithmetic mean

$$F_{t+1} = \frac{1}{t} \sum_{i=1}^{t} Y_i$$

Simple method but:

  - time series stationary in mean and variance
  - no seasonality

- Moving averages

$$F_{t+1} = \frac{1}{k} \sum_{i=t-k+1}^{t} Y_i$$

A similar idea to moving average as a component extraction method:

  - Smoothing: Trend-cycle component extracted as Average of observation around at instant t
  - Forecast: average of recent observations

2 issues to be considered

1. moving average length (which k?)
2. presence of seasonality and level shift

K choice

Lets think to one-step ahead forecasts $F(t+1)|t$

$$F_{t+1} = \frac{Y_t + Y_{t-1} + \ldots + Y_{t-k+1}}{k}$$

$$F_{t+2} = \frac{Y_{t+1} + Y_t + \ldots + Y_{t-k+2}}{k}$$

Comparing F_{t+1} e F_{t+2} we need to eliminate Y_{t-k+1} and add Y_{t+1}

$$F_{t+2} = F_{t+1} + \frac{1}{k}(Y_{t+1} - Y_{t-k+1})$$

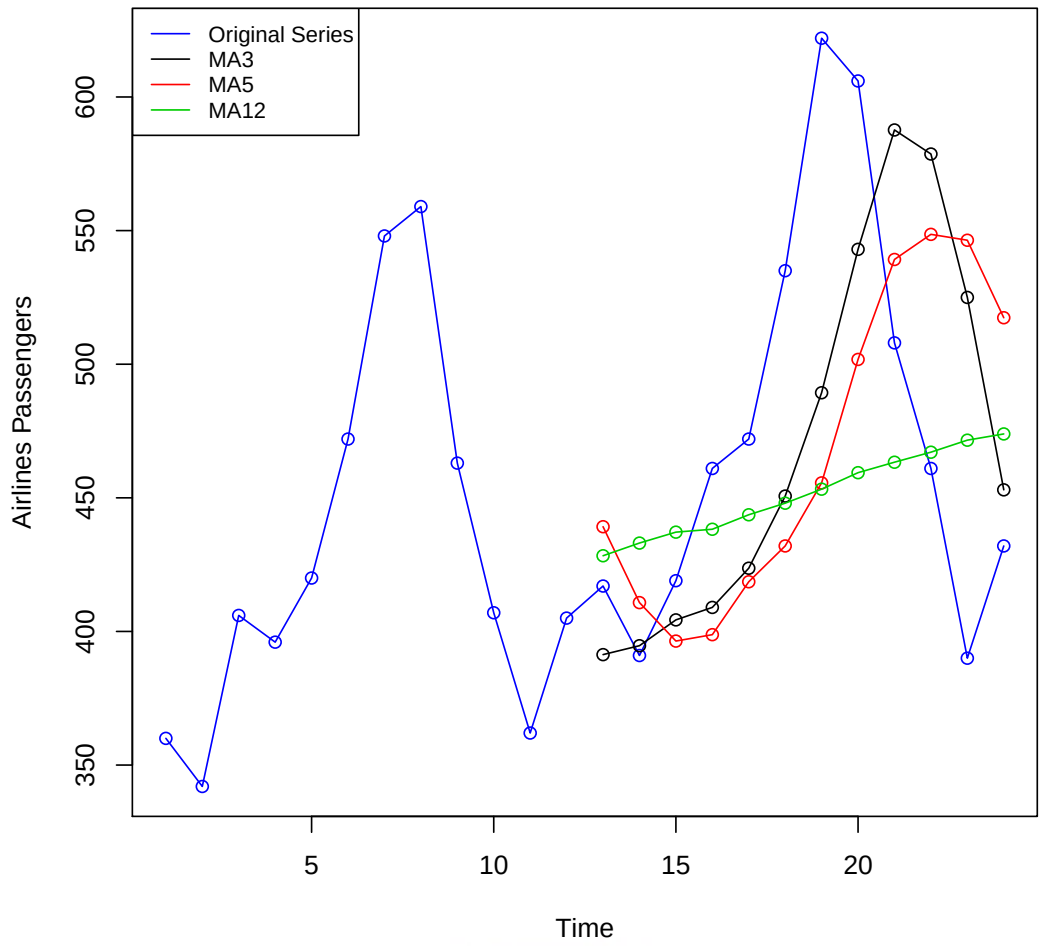
F_{t+2} is an adjustment F_{t+1} and k represents the weight in correction factor.

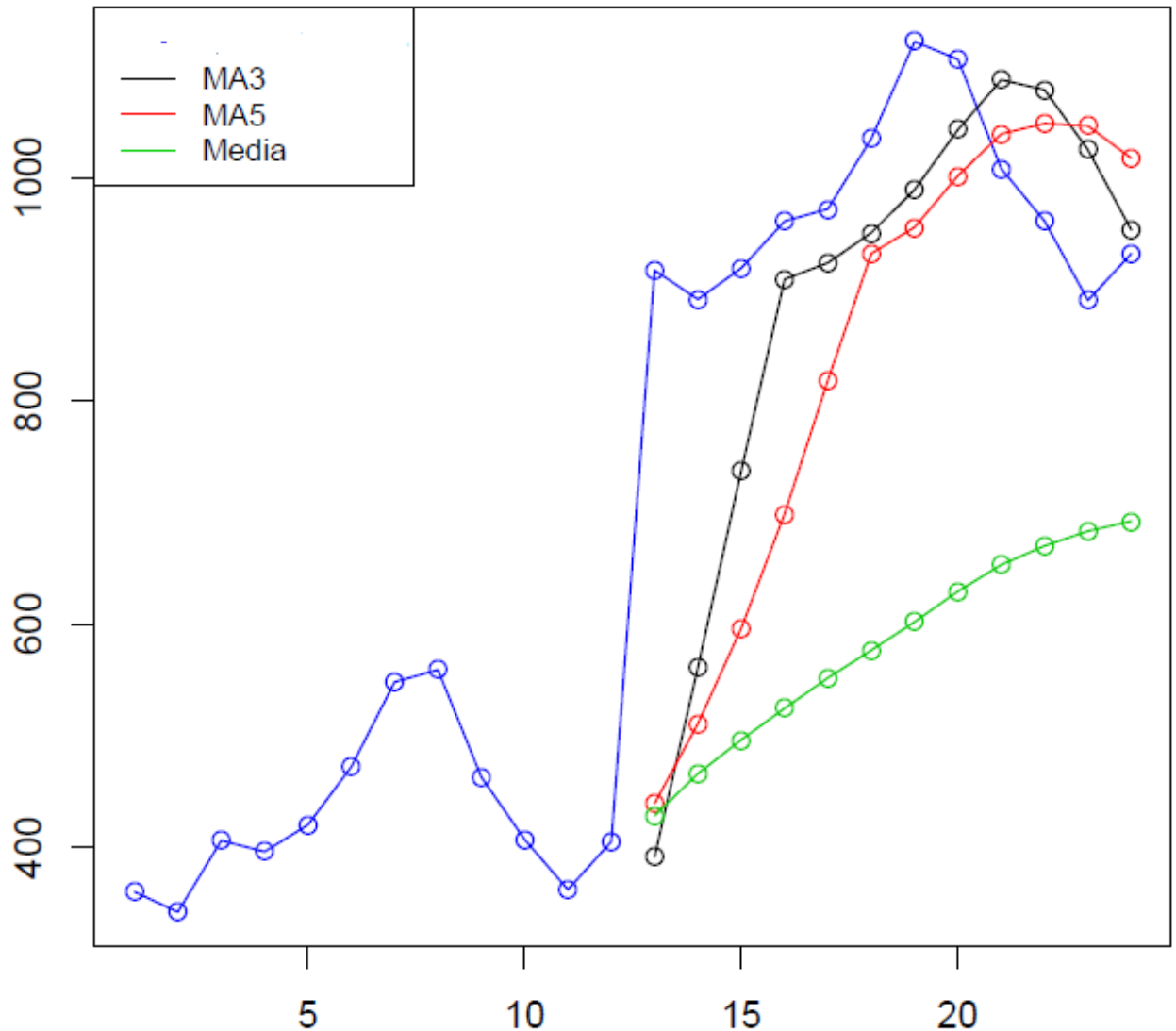
So:

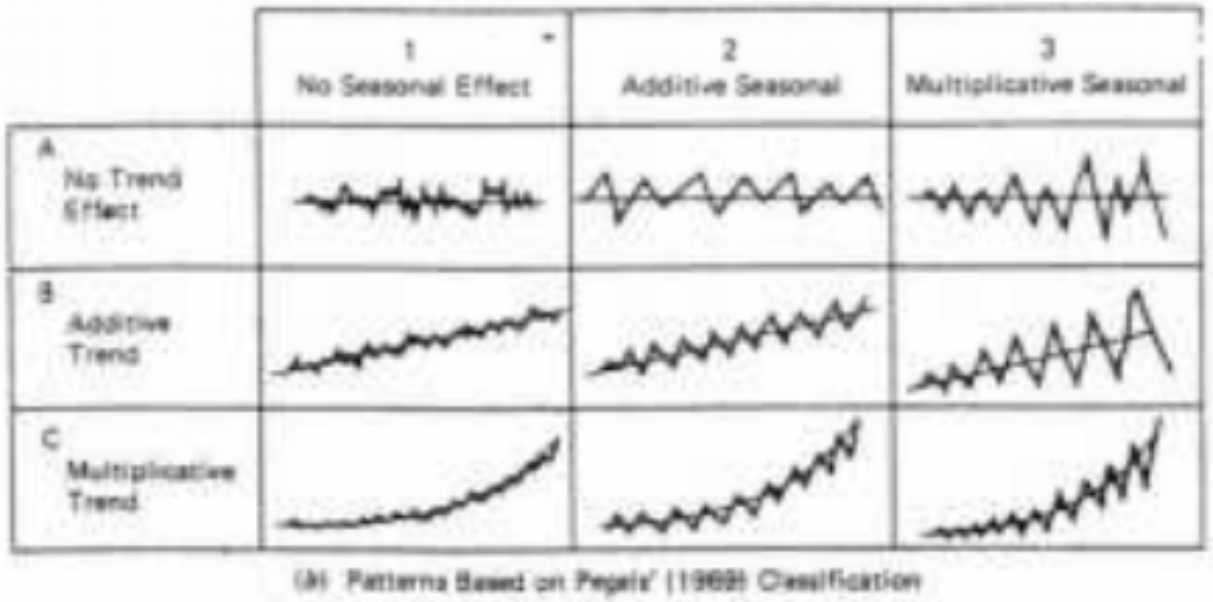
If k is big the adjustment is small and moving average with higher order have less variability

```
> nprev <- 12
> a10=AirPassengers[121:144]
> n <- length(a10) training = n-nprev
> ma3 <- ma5 <- ma12<- rep(NA,nprev)
> for(iprev in 1:nprev) {
  > ma3[iprev]  = mean(a10[(training-3+iprev):(training+iprev-1)])
  > ma5[iprev]  = mean(a10[(training-5+iprev):(training+iprev-1)])
  > ma12[iprev] = mean(a10[(training-12+iprev):(training+iprev-1)])}
>plot(1:n,a10,type="o",col=4,xlab="Time",ylab="Airlines")
>lines((training+1):n,ma3,type="o",col=1)
>lines((training+1):n,ma5,type="o",col=2)
>lines((training+1):n,ma12,type="o",col=3)
> legend("topleft",legend=c("Original series","MA3","MA5","MA(1)2")
+ ,col=c(4,1:3),lty=1,cex=0.8)
```

Different lengths




MA3
MA5
Media
1000
800
600
400
5
10
15
20



(A) Patterns Based on Pegels' (1969) Classification

Figure 4-1: *Patterns based on Pegels' (1969) classification.*

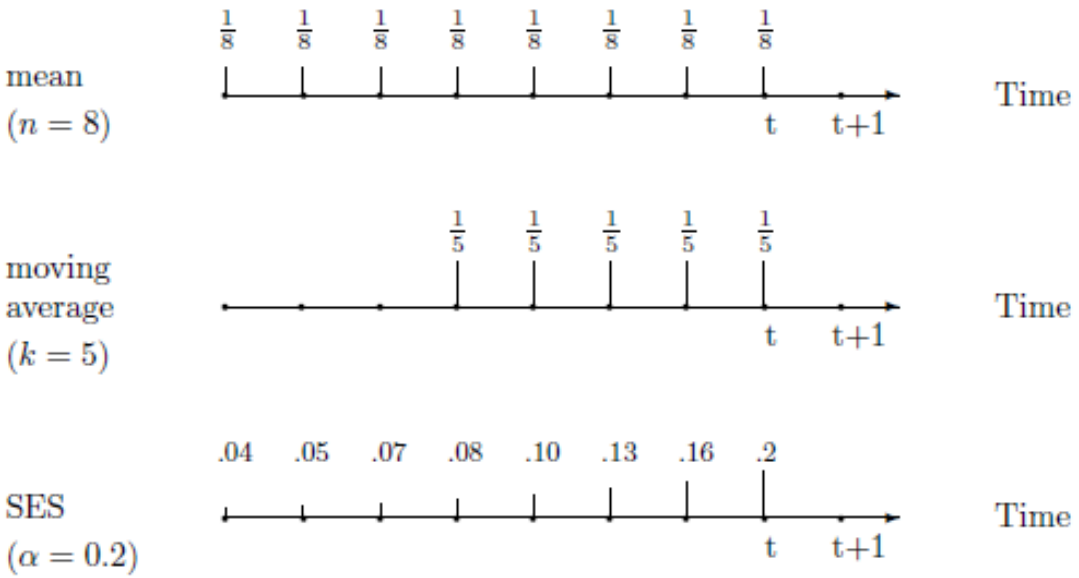
	Seasonal Component		
Trend Component	N (None)	A (Additive)	M (Multiplicative)
N (None)	(N,N)	(N,A)	(N,M)
A (Additive)	(A,N)	(A,A)	(A,M)
A_d (Additive damped)	(A_d,N)	(A_d,A)	(A_d,M)
M (Multiplicative)	(M,N)	(M,A)	(M,M)
M_d (Multiplicative damped)	(M_d,N)	(M_d,A)	(M_d,M)

(N,N)	=	simple exponential smoothing
(A,N)	=	Holts linear method
(M,N)	=	Exponential trend method
(A_d,N)	=	additive damped trend method
(M_d,N)	=	multiplicative damped trend method
(A,A)	=	additive Holt-Winters method
(A,M)	=	multiplicative Holt-Winters method
(A_d,M)	=	Holt-Winters damped method

Simple exponential smoothing

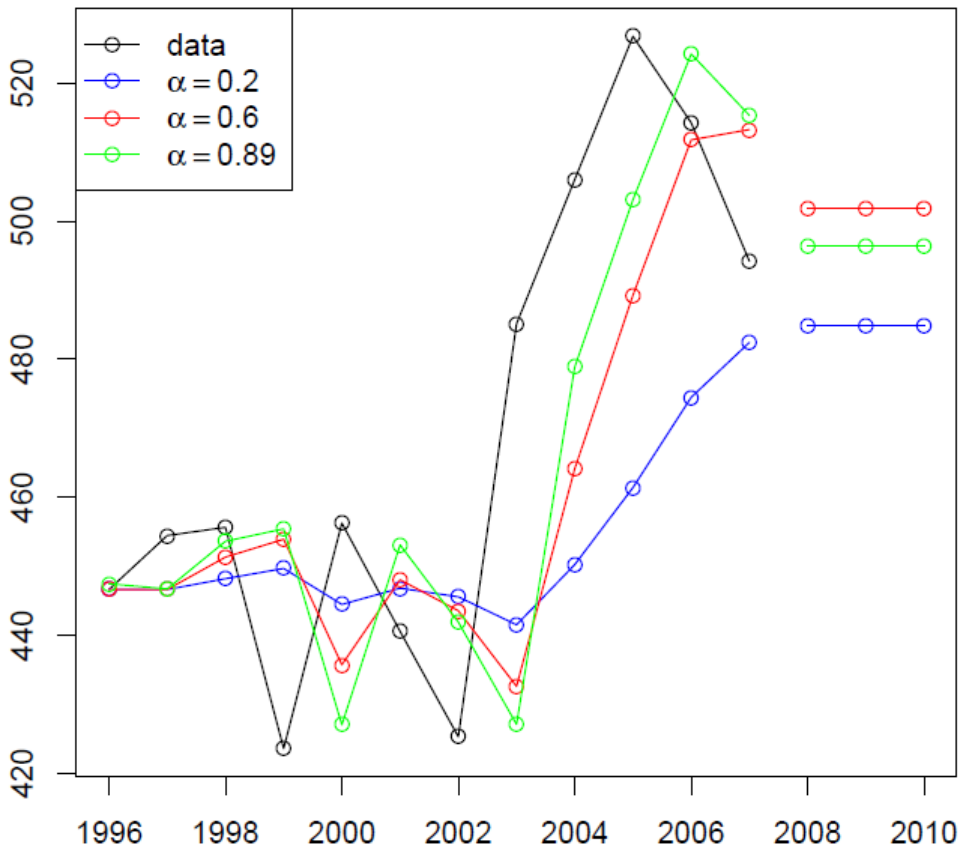
$$F_{t+1} = F_t + \alpha(Y_t - F_t) = \alpha Y_t + (1-\alpha)F_t$$

$$F_{t+1} = \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \ldots + (1-\alpha)^t F_1$$



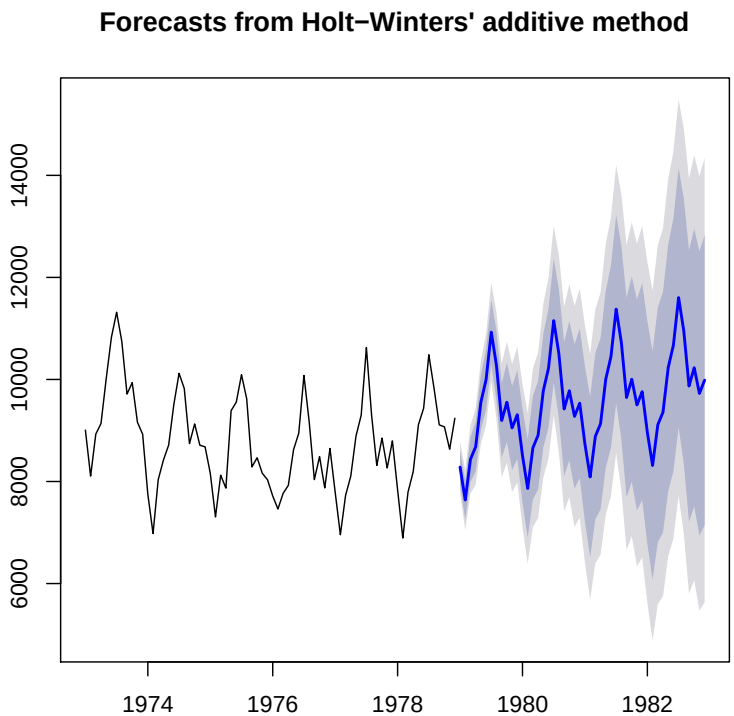
Trend	Seasonal: N	A	M
	Forecast equation / Smoothing equation		
N	$\hat{y}_{t+h\|t} = \ell_t$ $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = \ell_t s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t/\ell_{t-1}) + (1-\gamma)s_{t-m}$
A	$\hat{y}_{t+h\|t} = \ell_t + hb_t$ $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + hb_t + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = (\ell_t + hb_t)s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} + b_{t-1})) + (1-\gamma)s_{t-m}$
A_d	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t$ $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - \phi b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = (\ell_t + \phi_h b_t)s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} + \phi b_{t-1})) + (1-\gamma)s_{t-m}$
M	$\hat{y}_{t+h\|t} = \ell_t b_t^h$ $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t b_t^h + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = \ell_t b_t^h s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1}b_{t-1})) + (1-\gamma)s_{t-m}$
M_d	$\hat{y}_{t+h\|t} = \ell_t b_t^{\phi_h}$ $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}b_{t-1}^{\phi}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$	$\hat{y}_{t+h\|t} = \ell_t b_t^{\phi_h} + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}^{\phi}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$ $s_t = \gamma(y_t - \ell_{t-1}b_{t-1}^{\phi}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = \ell_t b_t^{\phi_h} s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}^{\phi}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$ $s_t = \gamma(y_t/(\ell_{t-1}b_{t-1}^{\phi})) + (1-\gamma)s_{t-m}$

```
fit1 = ses(oildata, alpha=0.2, initial="simple", h=3)
fit2 = ses(oildata, alpha=0.6, initial="simple", h=3)
fit3 = ses(oildata, h=3)
```



```
> summary(fit3)

Forecast method: Simple exponential smoothing

Model Information:
Simple exponential smoothing

Call:
 ses(x = oildata, h = 3)

  Smoothing parameters:
    alpha = 0.8921

  Initial states:
    l = 447.4808

  sigma:  25.1221

     AIC     AICc      BIC
111.1888 112.5221 112.1586

Error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE         ACF1
Training set 4.578526 25.12207 20.05797 0.8091703 4.252042 0.9245025 -0.03891839

Forecasts:
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2008       496.4923 464.2971 528.6876 447.2540 545.7307
2009       496.4923 453.3488 539.6359 430.5100 562.4747
2010       496.4923 444.6638 548.3208 417.2275 575.7572
```

```
deaths.fcast = hw(USAccDeaths,h=48)
plot(deaths.fcast)
```



Suppose to have an observed series $(y_1, y_2, \ldots, y_n)$ that is a realisation from the general process ARIMA (p,d,q)

$$\phi(L)\Delta^d y_t = \theta(L)\epsilon_t$$

and we wish to forecast a future value y_{T+h} (Granger and Newbold use n (now) and h (hence)). We use the linear filter representation

$$y_{t+h} = \epsilon_{t+h}\psi_1\epsilon_{t+h-1} + \ldots + \psi_{h+1}\epsilon_{t+1} + \psi_h\epsilon_t + \psi_{h+1}\epsilon_{t-h} + \ldots$$

where coefficient ψ are derived from

$$\Psi(L) = \theta^{-1}(L)\Delta^{-d}\theta(L)$$

the forecasts of y_{t+h} will be a linear combination of the past and present value of the observed realisation. One way is to regard the forecast as the conditional expectation (it is possible to demonstrate this result also using a quadratic cost function):

$$f_{t,h} = E(y_{t+h}|y_t, y_{t-1,\ldots}) = E((\epsilon_{t+h} + \psi_1\epsilon_{t+h-1} + \ldots + \psi_h\epsilon_n + \ldots)|x_n \ldots)$$

where $E(\epsilon_{t+j}$ is equal to 0 if $j > 0$ since future value have zero expectation while past value are known an equal to ϵ_{t+j} when $j \leq 0$:

$$f_{t,h} = \psi_h\epsilon_t + \psi_{h+1}\epsilon_{t-1} + \ldots$$

that is the minimum mean square error (MMSE)

forecast error is defined as:

$$e_{t,h} = y_{t+h} - f_{t,h} = \epsilon_{t+h} + \psi_1 \epsilon_{t+h-1} + \ldots + \psi_{h-1} \epsilon_{n+1}$$

so $f_{t,h}$ is an unbiased forecast with variance:

$$V(e_{t,h}) = \sigma^2 (1 + \psi_1^2 + \psi_2^2 + \ldots + \psi_{h-1}^2)$$

Suppose now to use an AR process. We know that:

$$\hat{y}_{t+1} = E(y_{t+1}|\Im_t)$$

this implies that $E(y_{t-j}|\Im_t) = y_{t-j}$ for each $j \geq 0$ and $E(\epsilon_{t+1}|\Im_t) = E(\epsilon_{t+1}) = 0$.
Forecast for y_{t+1} is:

$$\hat{y}_{t+1} = \phi_1 y_t + \ldots + \phi_p y_{t-p+1}$$

in an iterative way it is possible to define $\hat{y}_{t+2}$

Consider the following process AR(2):

$$y_t = 0.9y_{t-1} - 0.5y_{t-2} + \epsilon_t$$

and suppose that $y_{t-1} = 2$ and $y_t = 1$. It follows that the best forecast of y_{t+1} is:

$$\hat{y}_{t+1} = 0.9 \times 1 - 0.5 \times 2 = 0.1$$

and

$$\hat{y}_{t+2} = 0.9 \times (-0.1) - 0.5 \times 1 = -0.59$$

More generally, it is possible to define an impulse response function. Let y_t be an ARMA(2,1) process:

$$y_t = y_{t-1} - 0.5y_{t-1} + \epsilon_t + 0.75\epsilon_{t-1}$$

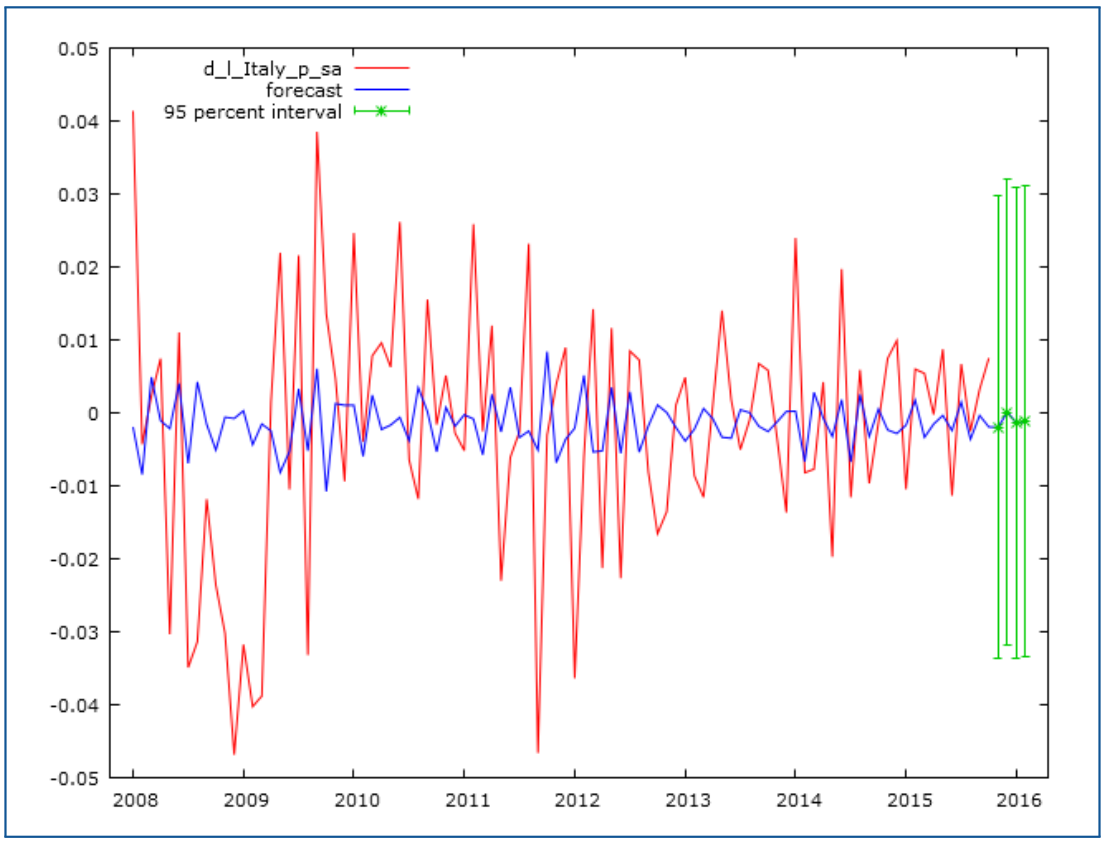
and suppose that at time t there is an expected event $\epsilon_t = 1$. What will be the effect on the process?

t	e_t	i_t
-2	0	0
-1	0	0
0	1	$i_{-1} - 0.5i_{-2} + e_0 + 0.75e_{-1} = 1$
1	0	$i_0 - 0.5i_{-1} + e_1 + 0.75e_0 = 1.75$
2	0	$i_1 - 0.5i_0 + e_2 + 0.75e_1 = 1.25$
3	0	$i_2 - 0.5i_1 + e_3 + 0.75e_2 = 0.375$
⋮	⋮	⋮

GOPA GOPA

European Commission

Example: 4 steps ahead of Italian Ipi (delta log of seasonal adjusted data) using an AR(2) model



GOPA GOPA

European Commission

Exercise 3
using real data

- check for stationarity
- check for transformation of the variables
- estimate an ARIMA model
- forecast the model

VECTOR AUTOREGRESSIVE MODELS

The extension of AR models into a multivariate framework leads to

- **V**ector **A**uto**R**egressive (**VAR**) models

VAR models became popular for economic analysis when Sims (1980) advocated them as alternatives to simultaneous equations models.

They are one of the most applied models in the empirical economics. VAR models are easy to use for forecasting and can also be applied for economic analysis.

Example: A Trivariate VAR(1)

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} + \begin{bmatrix} .5 & 0 & 0 \\ .1 & .1 & .3 \\ .0 & .2 & .3 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}$$

Example: A Bivariate VAR(2)

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix} + \begin{bmatrix} .5 & .7 \\ .9 & .2 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} .2 & .3 \\ .1 & .2 \end{bmatrix} \begin{bmatrix} y_{1t-2} \\ y_{2t-2} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

In a VAR model, every variable is allowed to depend linearly on its own history as well as on past values of the other variable in the VAR system, on deterministic variables, e.g. a constant or a deterministic trend term, and on an error term.

We start with the definition of Vector White Noise Process.

Definition (Vector White Noise Process)

A $(K \times 1)$ vector valued stochastic process $\{u_t = (u_{1t}, ..., u_{Kt})'; t \in \mathbb{Z}\}$ is said to be a K-dimensional white noise if

$$E(u_t) = 0_K$$
$$E(u_t u_t') = \Sigma_u$$
$$E(u_t u_s') = 0_{K \times K} \quad (t \neq s)$$

where covariance matrix Σ_u is assumed to be nonsingular if not otherwise is stated.

GOPA GOPA

European Commission

Definition (Vector Autoregressive Model)

Let $\{y_t = (y_{1t}, ..., y_{Kt})'; t \in \mathbb{Z}\}$ be a K-variate stochastic process. We say that the process $\{y_t; t \in \mathbb{Z}\}$ follows a vector autoregressive model of order p, denoted VAR(p) if

$$y_t = \nu + A_1 y_{t-1} + ... + A_p y_{t-p} + u_t, \quad t \in \mathbb{Z}$$

where

- p is a positive integer,
- A_i are fixed $(K \times K)$ coefficient matrices,
- $\nu = (\nu_1, ..., \nu_K)'$ is a fixed $(K \times 1)$ vector of intercept terms,
- $u_t = (u_{1t}, ..., u_{Kt})'$ is a K-dimensional white noise with nonsingular covariance matrix Σ_u.

Using the lag operator L, the VAR model can be written more compactly as

$$A(L)y_t = u_t$$

where $A(L) = I_K - A_1 L - \ldots - A_p L^p$ is a matrix polynomial in the lag operator L of order p.

Definition (Stability condition)

Let $\{y_t; t \in \mathbb{Z}\}$ be a K-dimensional process that follows a VAR(p)

$$y_t = \nu + A_1 y_{t-1} + ... + A_p y_{t-p} + u_t, \quad t \in \mathbb{Z}$$

We say that $\{y_t; t \in \mathbb{Z}\}$ is **stable** if

$$\det(I_K - A_1 z - ... - A_p z^p) \neq 0 \quad \text{for } |z| \leq 1.$$

that is, if all roots of the determinantal polynomial are outside the complex unit circle.

Example: Stable Bivariate VAR(2)

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} + \begin{bmatrix} .5 & .1 \\ .4 & .5 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ .25 & 0 \end{bmatrix} \begin{bmatrix} y_{1t-2} \\ y_{2t-2} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

We have that

$$\det\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} .5 & .1 \\ .4 & .5 \end{bmatrix} z - \begin{bmatrix} 0 & 0 \\ .25 & 0 \end{bmatrix} z^2 \right) = 1 - z + .21z^2 - .025z^3$$

The roots of this polynomial are

$$z_1 = 1.3, \quad z_2 = 3.55 + 4.26i \quad z_2 = 3.55 - 4.26i$$

The process is stable because all roots are outside the unit circle.

In order to understand what the stability assumption implicate, it is worth to visualize time series generated by stable VAR process and contrast it with realizations from unstable process.

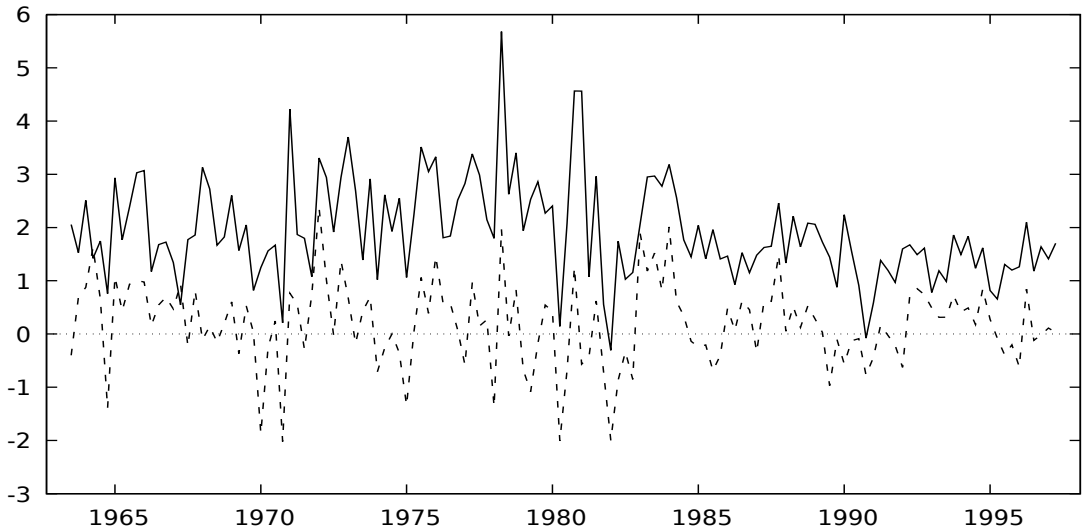


Figure 1: Bivariate time series generated by stable VAR processes

In Figure 1 a pair of time series generated by a stable bivariate VAR process is given, we note that they fluctuate around a constant mean and their variance does not change as their go along.

On the other hand, the pairs of series plotted in Figure 2 are generated by an unstable, bivariate processes. Clearly, the series in Figure 2 have a trend.

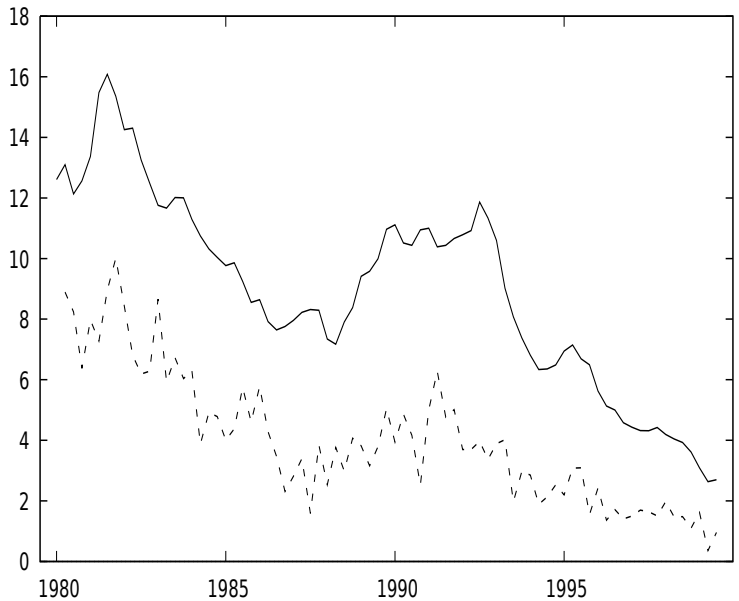


Figure 2: Bivariate time series generated by unstable processes

Definition (Stationary K-dimensional process)

Let $\{y_t; t \in \mathbb{Z}\}$ be a K-dimensional process. We say that $\{y_t; t \in \mathbb{Z}\}$ is stationary if

1. $E(y_t) = \mu \ \ \forall t \in \mathbb{Z}$
2. $E\left[(y_t - \mu)(y_{t-h} - \mu)'\right] = \Gamma_y(h) = \Gamma_y(-h)' \ \ \forall t \in \mathbb{Z}$ and $h = 0, 1, 2, ...$

In other words, a K-dimensional process $\{y_t; t \in \mathbb{Z}\}$ is stationary if its first and second moments are time invariant.

Theorem

Let $\{y_t; t \in \mathbb{Z}\}$ be a K-dimensional process that follows a VAR(p) model

$$y_t = \nu + A_1 y_{t-1} + ... + A_p y_{t-p} + u_t, \quad t \in \mathbb{Z}.$$

If $\{y_t; t \in \mathbb{Z}\}$ is stable, then $\{y_t; t \in \mathbb{Z}\}$ is stationary.

Stability implies stationarity

The converse of the Theorem is not true. An unstable VAR process is not necessarily nonstationary.

Again on the stability condition

$$y_{1t} = a_{10} + a_{11}Ly_{1t} + a_{12}Ly_{1t} + e_{1t}$$
$$y_{2t} = a_{20} + a_{21}Ly_{1t} + a_{22}Ly_{2t} + e_{2t}$$

or

$$(1 - a_{11}L)y_{1t} = a_{10} + a_{12}Ly_{1t} + e_{1t}$$
$$(1 - a_{22}L)y_{2t} = a_{20} + a_{21}Ly_{1t} + e_{2t}$$

So if we use this last equation

$$Ly_{2t} = \frac{L(a_{20} + a_{21}Ly_{1t} + e2t)}{(1 - a_{22}L)}$$

so that

$$(1 - a_{11}L)y_{1t} = a_{10} + a12L[\frac{L(a_{20} + a_{21}Ly_{1t} + e2t)}{(1 - a_{22}L)}] + e1t$$

Explicitly solving for y_{1t} gives

$$y_{1t} = \frac{a_{10}(1 - a_{22} + a_{12}a_{20} + (1 - a_{22}L)e_{1t} + a_{12}e_{2t-1}}{(1 - a_{11}L)(a_{22}L) - a_{12}a_{21}L^2}$$

In the same fashion this is the solution for y_{2t}

$$y_{2t} = \frac{a_{20}(1 - a_{11} + a_{21}a_{10} + (1 - a_{11}L)e_{2t} + a_{21}e_{1t-1}}{(1 - a_{11}L)(a_{22}L) - a_{12}a_{21}L^2}$$

Both equations have the same characteristic equation, so converge is related to the roots of the polynomial at the denominator.

GOPA GOPA

European Commission

ESTIMATION OF A VAR MODEL

Consider a bivariate VAR(1)

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

or

$$y_{1t} = \nu_1 + a_{11}y_{1t-1} + a_{12}y_{2t-1} + u_{1t}$$

$$y_{2t} = \nu_1 + a_{21}y_{1t-1} + a_{22}y_{2t-1} + u_{2t}$$

where $cov(u_{1t}, u_{2s}) = \sigma_{12}$ for $t = s$, 0 otherwise.

The model corresponds to 2 regressions with different dependent variables and identical explanatory variables. Further, we note all explanatory variables are predetermined variables and the error terms are assumed to be serially uncorrelated with constant variance.
Hence, we could estimate this model using the ordinary least squares (OLS) estimator computed separately from each equations.

OLS estimates are **consistent** and **asymptotically efficient**.

In general, the cooefficients of a K-dimensional VAR(p) model

$$y_t = \nu + A_1 y_{t-1} + ... + A_p y_{t-p} + u_t$$

can be estimated, without loss of estimation efficiency, by ordinary least squares. In the first equation, we have to run the regression

$$y_{1t} \text{ on } y_{1t-1}, \ldots, y_{Kt-1}, \ldots, y_{1t-p}, \ldots, y_{Kt-p},$$

in the second equation, we regress

$$y_{2t} \text{ on } y_{1t-1}, \ldots, y_{Kt-1}, \ldots, y_{1t-p}, \ldots, y_{Kt-p},$$

and so on.

VAR order selection

Before we can estimate a VAR model we must specify the order p. The most common approach for model order selection involves selecting a model order that minimizes one or more information criteria evaluated over a range of model orders.

Various lag length selection criteria (Akaike Information Criterion (AIC), Schwarz Information Criterion (BIC), Hannan-Quinn Criterion (HQ)) have been proposed in the literature (see Lutkepohl (2005]) for a detailed exposition of these criteria).

These measures are defined as:
Akaike Information Criterion (AIC):

$$AIC(p) = \ln\left|\tilde{\Sigma}\right| + \frac{2K^2p}{T}$$

where K is the number of variables in the system, T is the sample size, and $\tilde{\Sigma}$ is an estimate of the covariance matrix Σ.
Bayesian Information Criterion (BIC).

$$BIC(p) = \ln\left|\tilde{\Sigma}\right| + \frac{K^2p\ln T}{T}$$

Hannan-Quinn Criterion:

$$HQC(p) = \ln\left|\tilde{\Sigma}\right| + \frac{K^2p2\ln\ln T}{T}$$

The key difference between the criteria is how severely each penalizes increases in model order (the second term).

Lag selection: traditional framework, example

	constant only			constant and trend		
lags	AIC	BIC	HQC	AIC	BIC	HQC
1	4.405	4.632	4.497	4.338	4.621	4.453
2	4.008	**4.405**	4.169	3.984	**4.438**	4.168
3	3.880	4.447	4.110	3.860	4.484	4.114
4	3.914	4.651	4.214	3.896	4.690	4.218
5	**3.731**	4.638	**4.099**	**3.698**	4.662	**4.090**
6	3.744	4.821	4.181	3.710	4.844	4.170
7	3.763	5.010	4.269	3.728	5.032	4.257

So, possible choices are 2 and 5 (3 also attractive). Let's go for the safest choice (5).

Estimation In this case, we end up estimating

$$y_t = (\mu_0 + \mu_1 t) + \sum_{i=1}^{5} A_i y_{t-1} + u_t$$

via a battery of OLS regressions (no sweat).

Diagnostic Tests

Once a VAR model has been estimated, it is of pivotal interest to see whether the residuals obey the model's assumptions. That is, one should check for the absence of serial correlation and heteroscedasticity and see if the error process is normally distributed.

1. absence of serial correlation

Diagnostic Tests

Once a VAR model has been estimated, it is of pivotal interest to see whether the residuals obey the model's assumptions. That is, one should check for the absence of serial correlation and heteroscedasticity and see if the error process is normally distributed.

1. absence of serial correlation
2. absence of heteroscedasticity

Diagnostic Tests

Once a VAR model has been estimated, it is of pivotal interest to see whether the residuals obey the model's assumptions. That is, one should check for the absence of serial correlation and heteroscedasticity and see if the error process is normally distributed.

1. absence of serial correlation
2. absence of heteroscedasticity
3. error process normally distributed

A useful tool is the visual inspection of residuals. Formal autocorrelation checks are also nice — more later.

What do we do at this point?

Given the relatively low computational effort involved, vector autoregressive models are frequently used for macroeconomic forecasting purposes.

Forecasting

- The principle is simple: under quadratic loss, best predictor is conditional mean.

Forecasting

- The principle is simple: under quadratic loss, best predictor is conditional mean.
- If DGP is $A(L)y_t = \varepsilon_t$, then $E\left[A(L)y_t \middle| \mathcal{F}_{t-1}\right] = 0$, so $A(L)\hat{y}_t = 0$

Forecasting

- The principle is simple: under quadratic loss, best predictor is conditional mean.
- If DGP is $A(L)y_t = \varepsilon_t$, then $E\left[A(L)y_t \middle| \mathcal{F}_{t-1}\right] = 0$, so $A(L)\hat{y}_t = 0$
- therefore, $\hat{y}_{t+k} = \sum_{i=1}^{p} A_i \hat{y}_{t+k-1}$, where obviously $\hat{y}_{t+k-1} = y_{t+k-i}$ for $k \leq i$.

Advantages of VAR models

1. They are easy to estimate (computational burden is minimal).
2. They have good forecasting capabilities.
3. In a VAR system is very easy to test for Granger non-causality and for cointegration.

Problems with VARs

1. So many parameters. If there are K equations, one for each K variables and p lags of each of the variables in each equation, $(K + pK^2)$ parameters will have to be estimated.
2. VARs are a-theoretical, since they use little economic theory. Thus VARs cannot used to obtain economic policy prescriptions.

The concept of Granger causality was first introduced by Wiener (1956) and later reformulated and formalized by Granger (1969).

Granger causality has played a prominent role in economics and has many important uses in time series econometrics. Recently the notion of Granger causality has been applied in fields ranging from physics to physiology to climate science.

Granger, C. W. J., (1969). Investigating casual relations by econometric models and cross spectral methods, Econometrica, 37, 424-438.

GOPA GOPA

European Commission

An informal definition

Conceptually, the idea of Granger causality is very simple.

Consider two time series $y = \{y_t; t \in \mathbb{Z}\}$ and $x = \{x_t, t \in \mathbb{Z}\}$, we have that y causes x, with respect to a given information set I_t, which is assumed to contain at least x_{t-j}, y_{t-j}, $j \geq 0$, if, at time t, x_{t+1} can be better predicted by using present and past values of y than by not doing so, all other information in I_t (including the present and past of x) being used in either case.

The definition of causality proposed by Granger has stimulated considerable controversy among econometric theorists. According to several authors Granger causality is not causality in a deep sense of the word. Some authors have questioned use of the term 'cause' for Granger concept of causality; for example, some suggest that a better term might be 'predictability'.

Although Granger's concept is not equivalent to causality in the strict sense of the word, it can be considered a necessary condition for 'real' (physical) causality.

GOPA GOPA

European Commission

Definition

Let $x = \{x_t;\ t \in \mathbb{Z}\}$, $y = \{y_t;\ t \in \mathbb{Z}\}$ and $z = \{z_t;\ t \in \mathbb{Z}\}$ be three discrete-time stochastic processes, with finite second moments.

We denote by $H_x(t)$, $H_y(t)$ and $H_{xy}(t)$ the information sets generated, respectively, by subsets $\{x_s,\ s \leq t\}$, $\{y_s;\ s \leq t\}$ and $\{x_s, y_s;\ s \leq t\}$.

By $H_{xyz}(t)$ and $H_{xz}(t)$ information sets generated, respectively, by $\{x_s, y_s, z_s;\ s$ and $\{x_s, z_s;\ s \leq t\}$.

Finally, we denote by $P(x_{t+1}|H)$ the optimal prediction of x_{t+1} given the information set H.

A first formal definition of Granger non-causality may be given as follows:

Definition (Bivariate Granger non-causality)

y does not Granger cause x, with respect to $H_{xy}(t)$ (denoted $y \nrightarrow x|H_{xy}(t)$), iff

$$P(x_{t+1}|H_{xy}(t)) = P(x_{t+1}|H_x(t)) \ \ \forall t \in \mathbb{Z}.$$

We note that a rigorous definition of Granger causality requires the specification of

1. the stochastic process to be predicted, x;
2. the available information set, $H_{xy}(t)$;
3. the reduced information set, $H_x(t)$.

GOPA GOPA

European Commission

Definition (Trivariate Granger non-causality)

y does not Granger cause x, with respect to $H_{xyz}(t)$ (denoted $y \nrightarrow x|H_{xyz}(t)$), iff

$$P(x_{t+1}|H_{xyz}(t)) = P(x_{t+1}|H_{xz}(t)) \; \forall t \in \mathbb{Z}.$$

It is important to underline that these definition can be extended to the case where x, y and z are more than a one dimensional process.

Causality analysis in the Granger's sense is often conducted in a bivariate system (x, y). However, it is well known that Granger causal links are sensitive to the available information set used in the analysis.

Changing the available information set, for example by extending or reducing the number of time series in the study, may lead to different Granger causal links.

A causal statement is never intrinsic but always relative to a given information set.

Generally the causal structure of a subprocess of a multivariate stochastic process does not allow conclusions concerning the causal structure of the higher dimensional process. Causal statements made as if the universe were bivariate will not necessarily retain their validity when embedded in a higher dimensional setting.

Two important cases:

- **Non-causality due to omitted variables**: $y \nrightarrow x|H_{xy}(t)$ and $y \rightarrow x|H_{xyz}(t)$;
- **Spurious causality**: $y \rightarrow x|H_{xy}(t)$ and $y \nrightarrow x|H_{xyz}(t)$.

GOPA GOPA

European Commission

Granger-Noncausality in Bivariate AutoRegressive Models

Theorem 1. Let $(x_t, y_t)'$ be a bi-dimensional stochastic process that follows a VAR(p) model

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \nu_x \\ \nu_y \end{bmatrix} + \begin{bmatrix} a_{11,1} & a_{12,1} \\ a_{21,1} & a_{22,1} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \cdots +$$
$$+ \begin{bmatrix} a_{11,p} & a_{12,p} \\ a_{21,p} & a_{22,p} \end{bmatrix} \begin{bmatrix} x_{t-p} \\ y_{t-p} \end{bmatrix} + \begin{bmatrix} u_{xt} \\ u_{yt} \end{bmatrix}$$

y does not Granger-cause x, with respect to the information set $H_{xy}(t)$ if and only if

$$a_{12,i} = 0 \quad \text{for } i = 1, ..., p.$$

This theorem implies that Granger-noncausality can be checked easily by looking at the VAR representation of the system.

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \nu_x \\ \nu_y \end{bmatrix} + \begin{bmatrix} .7 & 0 \\ .3 & .5 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} .1 & 0 \\ .2 & .2 \end{bmatrix} \begin{bmatrix} x_{t-2} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} u_{xt} \\ u_{yt} \end{bmatrix}$$

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \nu_x \\ \nu_y \end{bmatrix} + \begin{bmatrix} .7 & 0 \\ .3 & .5 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} .1 & 0.3 \\ .2 & .2 \end{bmatrix} \begin{bmatrix} x_{t-2} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} u_{xt} \\ u_{yt} \end{bmatrix}$$

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \nu_x \\ \nu_y \end{bmatrix} + \begin{bmatrix} .7 & .2 \\ .3 & .5 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} .1 & .5 \\ .2 & .2 \end{bmatrix} \begin{bmatrix} x_{t-2} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} u_{xt} \\ u_{yt} \end{bmatrix}$$

Here we assume that the our VAR is stable. In this system, we can checking causality of y for x by testing

$$H_0 : a_{12,i} = 0 \quad \text{for } i = 1, ..., p.$$

Posing
$\mathbf{\Phi} = [\nu, A_1, ..., A_p]$ a $(2 \times (2p + 1))$ matrix,
$\beta = \text{vec}(\mathbf{\Phi})$ a $((4p + 2) \times 1)$ vector.
Let $\mathbf{R}$ be a matrix such that $\mathbf{R}\beta = [a_{12,1}, \ldots, a_{12,p}]'$.
The null becomes

$$H_0 : \mathbf{R}\beta = \mathbf{0}.$$

To test

$$H_0 : \mathbf{R}\beta = \mathbf{0}.$$

we use the Wald statistic

$$W = T\hat{\beta}'\mathbf{R}' \left[\mathbf{R}V(\hat{\beta})\mathbf{R}'\right] \mathbf{R}\hat{\beta}$$

where $\hat{\beta}$ is a consistent estimator of β and $V(\hat{\beta})$ is a consistent estimator of the asymptotic variance-covariance matrix of $\sqrt{(T)}(\hat{\beta} - \beta)$.
Given appropriate conditions, W is asymptotically distributed as a $\chi^2(p)$ random variable under H_0.

We reject H_0 if $W > \chi^2_\alpha$

GOPA GOPA

European Commission

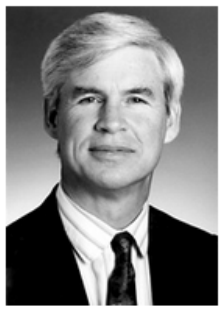
COINTEGRATION

The notion of cointegration has become one of the more important concepts in time series since the papers by Graner (1983) and Engle and Granger (1987).

GOPA GOPA

European Commission

The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2003



Robert F. Engle III
Prize share: 1/2

Clive W.J. Granger
Prize share: 1/2

The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2003 was divided equally between Robert F. Engle III *"for methods of analyzing economic time series with time-varying volatility (ARCH)"* and Clive W.J. Granger *"for methods of analyzing economic time series with common trends (cointegration)"*.

GOPA GOPA

European Commission

Most macroeconomic time series follow a stochastic trend, so that a temporary disturbance in, say, GDP has a long-lasting effect. These time series are called **nonstationary**; they differ from stationary series which do not grow over time, but fluctuate around a given value.

- Clive Granger demonstrated that the statistical methods used for stationary time series could yield wholly **misleading results when applied to the analysis of nonstationary data**
- His significant discovery was that specific combinations of nonstationary time series may exhibit stationarity, thereby allowing for correct statistical inference
- Granger called this phenomenon cointegration.

- Granger has shown that macroeconomic models containing nonstationary stochastic variables can be constructed in such a way that the results are both statistically sound and economically meaningful.
- His work has also provided the underpinnings for modeling with rich dynamics among interrelated economic variables. Cointegration concept has radically changed the way empirical models of macroeconomic relationships are formulated today.

GOPA GOPA

European Commission

Cointegration

If two series, y_t and x_t, are both I(1) and there exists a real number β such that $z_t = y_t - \beta x_t$ is I(0), we say that y_t and x_t are **cointegrated**. The vector (1, -β)' is said the **cointegrating vector**.

The notion of cointegration is related to the concept of a long-run equilibrium. It clear that cointegration reflects the presence of a a long-run equilibrium relationship between y_t and x_t.

If y_t and x_t are not cointegrated and, consequently, u_t is I(1), the equilibrium error u_t can wander widely and zero-crossings would be very rare. Under such circumstances, it does not make sense to refer to $y_t^* = \beta x_t$ a long-run equilibrium.

Consequently, the presence of a cointegration can be interpreted as the presence of a long-run equilibrium relationship.

Economic theory often suggests that certain pairs of economic or financial variables should be linked by a long-run economic relationship.

- The permanent income hypothesis (PIH) implies cointegration between consumption and income.
- Money demand models imply cointegration between money, nominal income, prices, and interest rates.
- The Fisher equation implies cointegration between nominal interest rates and inflation.
- Growth theory models imply cointegration between income, consumption, and investment.

Example of cointegration:

$$\begin{cases} x_{1t} = x_{1t-1} + \varepsilon_t \\ x_{2t} = x_{1t} + u_t \end{cases}$$

where ε_t and u_t are stationary.

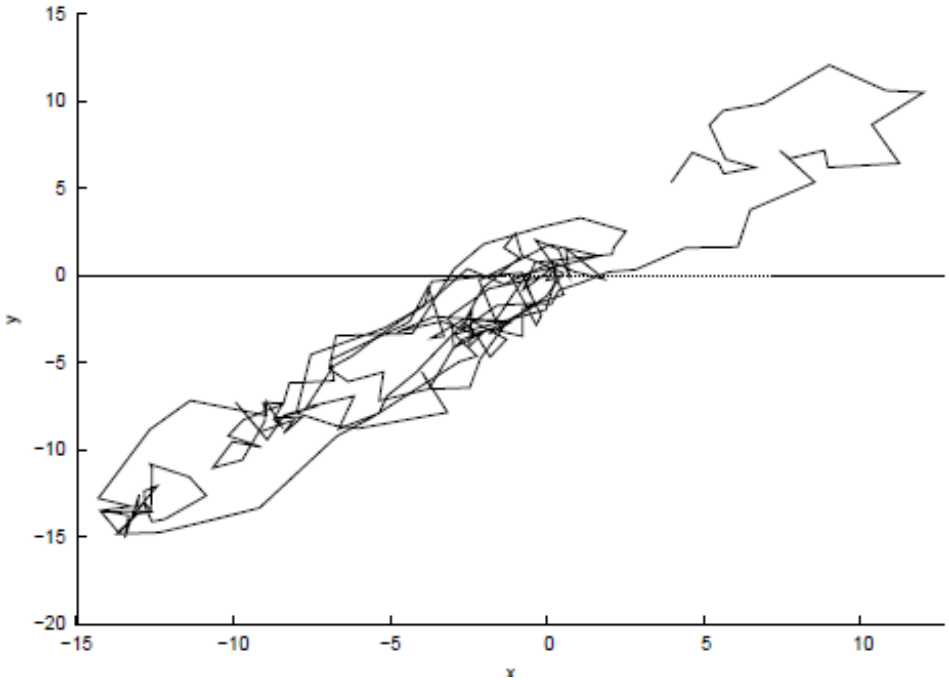
Clearly

- x_{1t} is $I(1)$
- x_{2t} is $I(1)$ too
- but $z_t = x_{2t} - x_{1t} = u_t$ is $I(0)$

in VAR form

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \omega_t \end{bmatrix}$$

where $\omega_t = \varepsilon_t + u_t$.


15
10
5
0
−5
−10
−15
−20
y
−15
−10
−5
0
5
10
x

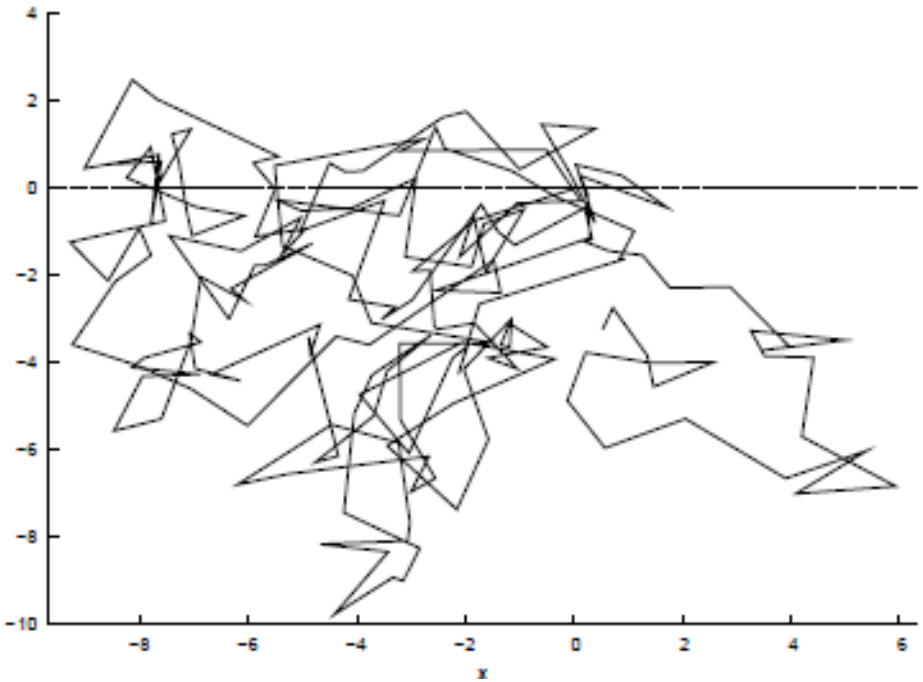
in VECM form

$$
\begin{aligned}
\begin{bmatrix} \Delta x_{1,t} \\ \Delta x_{2,t} \end{bmatrix} &= \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \omega_t \end{bmatrix} = \\
&= \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \omega_t \end{bmatrix} = \\
&= \begin{bmatrix} 0 \\ 1 \end{bmatrix} z_{t-1} + \begin{bmatrix} \varepsilon_t \\ \omega_t \end{bmatrix}
\end{aligned}
$$

GOPA GOPA

European Commission

random walk



Cointegration with more than two variables

Let $y_t = (y_{1t}, ..., y_{nt})'$ denote an $(n \times 1)$ vector of I(1) time series. y_t is cointegrated if there exists an $(n \times 1)$ vector $\beta = (\beta_1, ..., \beta_n)'$ such that

$$\beta' y_t = \beta_1 y_{1t} + ... + \beta_n y_{nt} \sim I(0)$$

In words, the nonstationary time series in y_t are cointegrated if there is a linear combination of them that is stationary or I(0).

Normalization. The cointegration vector β is not unique since for any scalar c the linear combination $c\beta' y_t = \beta^{*\prime} y_t \sim I(0)$. Hence, some normalization assumption is required to uniquely identify β. A typical normalization is

$$\beta = (1, -\beta_2, ..., -\beta_n)'$$

so that the cointegration relationship may be expressed as

$$\beta' y_t = y_{1t} - \beta_2 y_{2t} - ... - \beta_n y_{nt} \sim I(0)$$

or

$$y_{1t} = \beta_2 y_{2t} - ... - \beta_n y_{nt} + u_t$$

where $u_t \sim I(0)$.

Multiple Cointegrating Relationships. If the $(n \times 1)$ vector $y_t = (y_{1t}, ..., y_{nt})'$ is cointegrated there may be $0 < r < n$ linearly independent cointegrating vectors. For example, let $n = 3$ and suppose there are $r = 2$ cointegrating vectors $\beta_1 = (\beta_{11}, \beta_{12}, \beta_{13})'$ and $\beta_2 = (\beta_{21}, \beta_{22}, \beta_{23})'$. Then

$$\beta_1' y_t = \beta_{11} y_{1t} + \beta_{12} y_{2t} + \beta_{13} y_{3t} \sim I(0)$$

and

$$\beta_2' y_t = \beta_{21} y_{1t} + \beta_{22} y_{2t} + \beta_{23} y_{3t} \sim I(0)$$

.

The vectors β_1 and β_2 form a basis for the space of cointegrating vectors. Any linear combination of β_1 and β_2, e.g. $\beta_3 = c_1 \beta_1 + c_2 \beta_2$ where c_1 and c_2 are constants, is also a cointegrating vector.

Vector Error Correction Model (VECM). Consider a VAR(p) model

$$y_t = A_1 y_{t-1} + ... + A_p y_{t-p} + u_t$$

For the case with $p = 3$ we can write this as

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + A_3 y_{t-3} + u_t$$

or

$$A(L) y_t = \left(I - A_1 L - A_2 L^2 - A_3 L^3 \right) y_t = u_t$$

now rewrite this by re-expressing y_s as y_{t-1} and the appropriate deltas. That is

$$\begin{aligned} y_t &= y_{t-1} + \Delta y_t \\ y_{t-1} &= y_{t-1} \\ y_{t-2} &= y_{t-1} - \Delta y_{t-1} \\ y_{t-3} &= y_{t-1} - (\Delta y_{t-1} + \Delta y_{t-2}) \end{aligned}$$

Now substitute:

$$y_{t-1}+\Delta y_t = A_1 y_{t-1}+A_2\left(y_{t-1}-\Delta y_{t-1}\right)+A_3\left(y_{t-1}-\Delta y_{t-1}-\Delta y_{t-2}\right)+u_t$$

and re-arrange:

$$\Delta y_t = -(I - A_1 - A_2 - A_3)y_{t-1} - (A_2 + A_3)\Delta y_{t-1} - A_3\Delta y_{t-2} + u_t$$

and rename the parameters:

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \Gamma_2 \Delta y_{t-2} + u_t$$

note that $\Pi = -A(1)$.

Similarly, we can write for general values of p that

$$\Delta y_t = \Pi y_{t-1} + \alpha\beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + \ldots + \Gamma_{p-1}\Delta y_{t-p+1} + u_t$$

The matrix Π is called the long-run impact matrix and Γ_k are the short run impact matrices.

We remember that

$$\Pi = -A(1) = -(I - A_1 - ... - A_p)$$

If the VAR(p) process has unit roots then from

$$\det(I - A_1 z - ... - A_p z^p) = 0$$

it is clear that Π is a singular matrix.

If Π is singular then it has reduced rank; that is $\text{rank}(\Pi) = r < n$. There are two cases to consider:

1. $\text{rank}(\Pi) = 0$. This implies that $\Pi = 0$ and y_t is I(1) and not cointegrated. The VECM reduces to a VAR(p-1) in first differences

$$\Delta y_t = \alpha\beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + ... + \Gamma_{p-1} \Delta y_{t-p+1} + u_t$$

2. $0 < \text{rank}(\Pi) = r < n$. This implies that y_t is cointegrated with r linearly independent cointegrating vectors.

Some algebra: if Π is singular but not zero (that is $0 < \text{rank}(\Pi) = r < n$), then there are two $n \times r$ matrices α and β (with $0 < r < n$) such that $\Pi = \alpha\beta'$.
Therefore,

$$\Delta y_t = \alpha\beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + ... + \Gamma_{p-1} \Delta y_{t-p+1} + u_t$$

This model is referred to as a **vector error-correction model** (VECM).

Since $y_t \sim I(1)$, we have that $\Delta y_t \sim I(0)$. If the left-hand side is $I(0)$, so must be the right-hand side. This implies that $\beta' y_{t-1}$ must be $I(0)$ too. We can conclude that y_t is cointegrated with r linearly independent cointegrating vectors.

Note that the existence of n cointegrating relationships is impossible: if n independent linear combinations produce stationary series, all n variables themselves must be stationary.

GOPA GOPA

European Commission

Testing for Cointegration

We have seen that the rank of Π is crucial in determining the number of cointegration relationships.

- If $\text{rank}(\Pi) = 0$, then no cointegrating relationship exists.
- If $\text{rank}(\Pi) = n$, then any linear combination is stationary.
- If $0 < \text{rank}(\Pi) = r < n$, then multiple cointegrating relationships exist.

GOPA GOPA

European Commission

Testing for Cointegration: Johansen Approach

Testing for the number of significant cointegrating relationships using Johansen Approach involves checking the significance of the characteristic roots of Π.

Using the property that the rank of a matrix is equal to the number of characteristic roots that differer from zero, Johansen proposes two test for cointegration; one based upon the trace and one on the maximum eigenvalue.

The trace test. Let us denote the (theoretical) eigenvalues of this matrix in decreasing order as $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$. If there are r cointegrating relationships (and Π has rank r) it must be the case that $\log(1-\lambda_j) = 0$ for the smallest $n-r$ eigenvalues, that is, for $j = r+1, r+2, ..., n$.

We can use the (estimated) eigenvalues, say $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq ... \geq \hat{\lambda}_n$, to test hypotheses about the rank of Π. For example, the hypothesis $H_0 : r \leq r_0$ versus the alternative $H_1 : r_0 < r \leq n$, can be tested using the statistic

$$\lambda_{trace}(r_0) = -T \sum_{j=r_0+1}^{n} \log(1-\hat{\lambda}_j)$$

The trace test checks whether the smallest $n-r_0$ eigenvalues are significantly different from zero.

The eigenvalue test. We can test $H_0 : r \leq r_0$ versus the more restrictive alternative $H_1 : r = r_0 + 1$ using

$$\lambda_{max}(r_0) = -T\log(1 - \hat{\lambda}_{r_0+1}).$$

This alternative test is called the **maximum eigenvalue test**, as it is based on the estimated $(r_0 + 1)$th largest eigenvalue.

The two tests described here are actually likelihood ratio tests, but do not have the usual Chi-squared distributions. Instead, the appropriate distributions are multivariate extensions of the Dickey-Fuller distributions.

GOPA GOPA

European Commission

Procedure for Determining the Number of Cointegrating Vectors

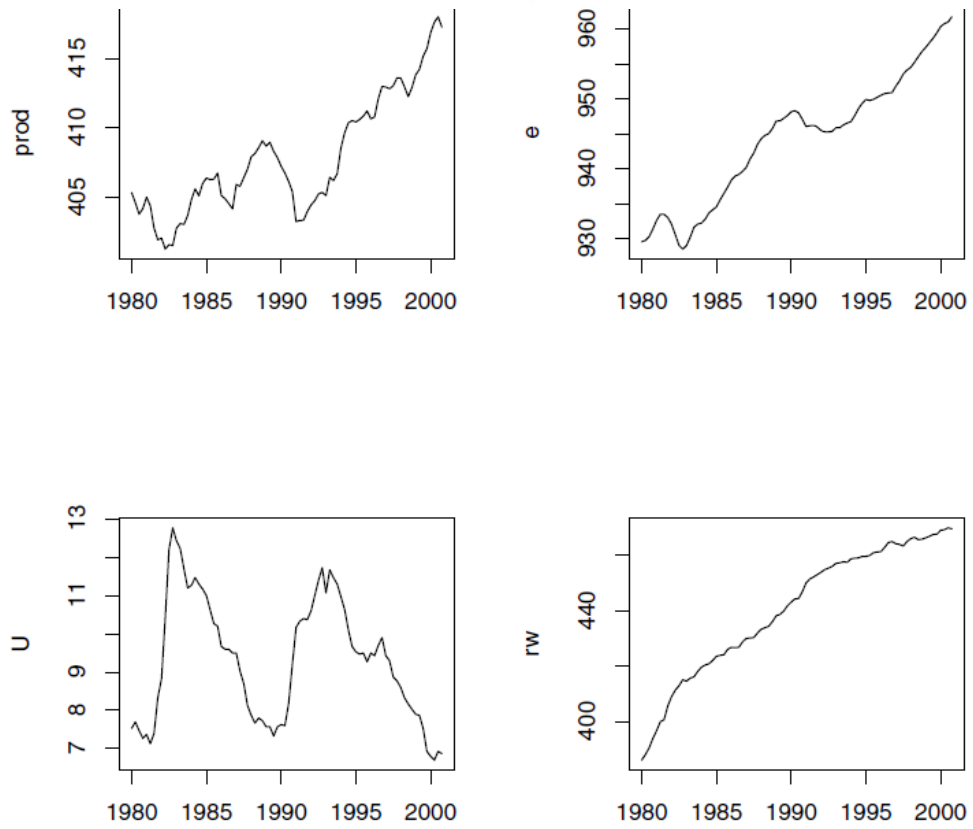
Johansen proposes a sequential testing procedure that consistently determines the number of cointegrating vectors.
First test $H_0 : r_0 = 0$ against $H_1 : r_0 > 0$. If this null is not rejected then it is concluded that there are no cointegrating vectors among the n variables in y_t. If $H_0 : r_0 = 0$ is rejected then it is concluded that there is at least one cointegrating vector and proceed to test $H_0 : r_0 = 1$ against $H_1 : r_0 > 1$. If this null is not rejected then it is concluded that there is only one cointegrating vector. If the null is rejected then it is concluded that there is at least two cointegrating vectors.
The sequential procedure is continued until the null is not rejected.

GOPA GOPA

European Commission

Example on Canadian data

- labour productivity as the log difference of GDP and employment - **prod**
- log of employment **e**
- log of unemployment **U**
- real wages **rw**
- source OECD database 1.1980 4-2004


prod
405
410
415
1980
1985
1990
1995
2000
e
930
940
950
960
1980
1985
1990
1995
2000
U
7
8
9
11
13
1980
1985
1990
1995
2000
rw
400
440
1980
1985
1990
1995
2000

GOPA GOPA

European Commission

Variable	Deterministic Terms	Lags	Test Value	Critical Value		
				1%	5%	10%
$prod$	constant, trend	2	−1.99	−4.04	−3.45	−3
$\Delta prod$	constant	1	−5.16	−3.51	−2.89	−2
e	constant, trend	2	−1.91	−4.04	−3.45	−3
Δe	constant	1	−4.51	−3.51	−2.89	−2
U	constant	1	−2.22	−3.51	−2.89	−2
ΔU		0	−4.75	−2.6	−1.95	−1
rw	constant, trend	4	−2.06	−4.04	−3.45	−3
Δrw	constant	3	−2.62	−3.51	−2.89	−2
Δrw	constant	0	−5.6	−3.51	−2.89	−2

GOPA GOPA

European Commission

Lag Order	AIC(n)	HQ(n)	SC(n)	FPE(n)
$p = 1$	-6.2726	-5.9784	-5.5366	0.0019
$p = 2$	-6.6367	-6.1464	-5.4100	0.0013
$p = 3$	-6.7712	-6.0848	-5.0538	0.0012
$p = 4$	-6.6346	-5.7522	-4.4265	0.0014
$p = 5$	-6.3981	-5.3196	-3.6994	0.0018
$p = 6$	-6.3077	-5.0331	-3.1183	0.0020
$p = 7$	-6.0707	-4.6000	-2.3906	0.0028
$p = 8$	-6.0616	-4.3947	-1.8908	0.0031

$\mathcal{H}_0$	Test Statistics		Critical Values		
	$p = 3$	$p = 2$	90%	95%	99%
$r = 0$	84.92	86.12	59.14	62.99	70.05
$r = 1$	36.42	37.33	39.06	42.44	48.45
$r = 2$	18.72	15.65	22.76	25.32	30.45
$r = 3$	3.85	4.10	10.49	12.25	16.26