Quality Data Analysis
4- Time series modeling via ARIMA

Bianca Maria Colosimo – biancamaria.colosimo@polimi.it

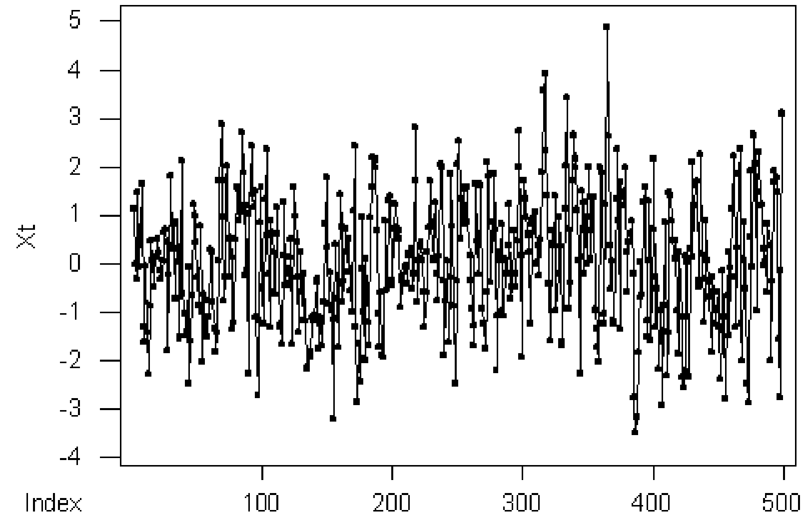
Reference:

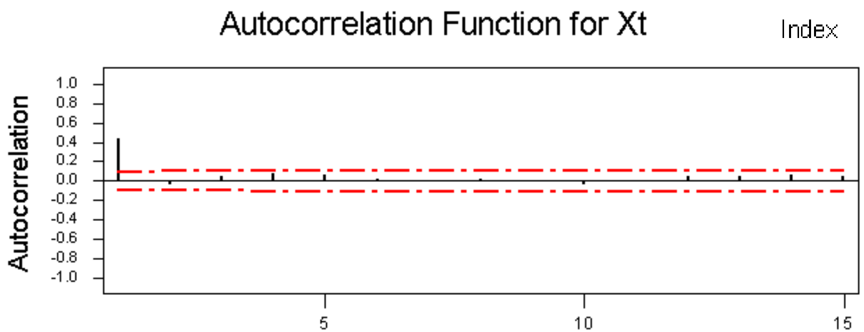
“Statistical Process Adjustment for Quality Control”, Del Castillo – Wiley

“Time Series Analysis – 3rd edition”, Box Jenkins Reinsel – Prentice Hall

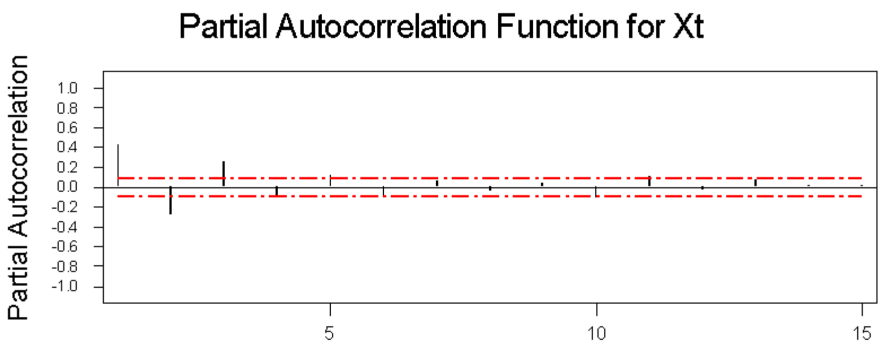
Why ARIMA?

Example





Example: 500 consecutive observations from a productive process



? Identification ?

More general models: ARIMA

Time series and stationarity

Let $\{X_t\}$ be a discrete time series, then:

The time series is ***strictly (or strongly) stationary*** if its properties do not depend on modifications of the time origin

I.e., the joint distribution of : $X_{t_1}, X_{t_2}, X_{t_3}, \ldots X_{t_m}$

coincides with the joint distribution of $X_{t_1+k}, X_{t_2+k}, X_{t_3+k}, \ldots X_{t_m+k} \quad \forall k$

We refer to ***weak stationarity of order f*** if all the moments of the series up to order *f* only depend on the time difference between the time series data

E.g.: stationarity of 2nd order: $\begin{cases} E(X_t) = \mu \\ Cov(X_t, X_{t-k}) = \gamma_k \end{cases} \quad \forall t = 1,2,\ldots$

Remind: $Cov(X_t, X_{t-k}) = \gamma_k = E[(X_t - \mu)(X_{t-k} - \mu)] \quad k = 0, \pm 1, \pm 2, \ldots$

ARMA(p,q)

$$\varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2)$$

General mode for stochastic model (linear)

$$X_t = \xi + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Most of stationary processes can be modeled by including:
- An autoregressive term of degree p - AR(p) term
- A 'moving average' term that links the observation at time t to previous q random errors - MA(q) term

Let's study the details

Model AR(1)
(Markov process)

$$X_t = \xi + \phi_1 X_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim NID(0, \sigma^2)$$

For a stationary process: $E(X_t) = \mu$

$$E(X_t) = \xi + \phi_1 E(X_{t-1}) + 0 \Rightarrow (1-\phi_1)\mu = \xi$$

$$X_t = (1-\phi_1)\mu + \phi_1 X_{t-1} + \varepsilon_t$$

$$X_t - \mu = \phi_1 (X_{t-1} - \mu) + \varepsilon_t \qquad \widetilde{X}_t \equiv X_t - \mu = \phi_1 (\widetilde{X}_{t-1}) + \varepsilon_t$$

Let’s introduce the backward shift (backshift) operator $B : BX_t = X_{t-1}$

$$\widetilde{X}_t = \phi_1 (\widetilde{X}_{t-1}) + \varepsilon_t = \phi_1 (B\widetilde{X}_t) + \varepsilon_t \Rightarrow \boxed{(1-\phi_1 B)\widetilde{X}_t = \varepsilon_t}$$

Generally speaking: AR(p) model:

$$X_t = \xi + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \varepsilon_t$$

Analogously, for a stationary process: $E(X_t) = \mu$

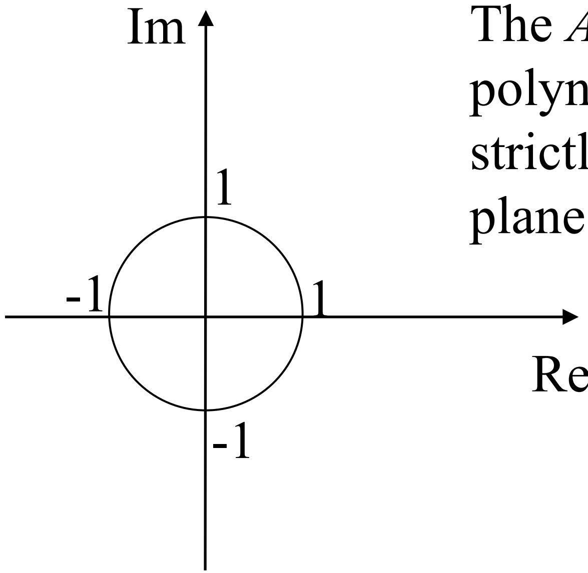
$$(1 - \sum_{i=1}^{p} \phi_i)\mu = \xi \Rightarrow \widetilde{X}_t \equiv X_t - \mu = \sum_{i=1}^{p} \phi_i \widetilde{X}_{t-i} + \varepsilon_t$$

Backshift operator $B : X_{t-2} = BX_{t-1} = B(BX_t) = B^2 X_t$

$$\widetilde{X}_t = \sum_{i=1}^{p} \phi_i B^i \widetilde{X}_t + \varepsilon_t \Rightarrow \boxed{A(B)\widetilde{X}_t = \varepsilon_t}$$

where $A(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$

Stationarity



The *AR(p)* process is stationary if and only if the polynomial *A(B)* is stable, i.e., all of its roots lie strictly outside the unit circle in the complex plane

E.g.: $AR(1): (1-\phi_1 B)\widetilde{X}_t = \varepsilon_t$

$$A(B) = 1-\phi_1 B = 0 \Rightarrow B = \frac{1}{\phi_1} \quad \begin{cases} |\phi_1| < 1 & \text{stationary} \\ |\phi_1| \geq 1 & \text{non stationary} \end{cases}$$

The Jury's test is used to test the stability of AR(p) processes

Some intuitive remarks

$$\tilde{X}_t = \phi_1(\tilde{X}_{t-1}) + \varepsilon_t = \phi_1(\phi_1(\tilde{X}_{t-2}) + \varepsilon_{t-1}) + \varepsilon_t = \phi_1^2 \tilde{X}_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t =$$
$$= \phi_1^2 \left(\phi_1(\tilde{X}_{t-3}) + \varepsilon_{t-2}\right) + \phi_1 \varepsilon_{t-1} + \varepsilon_t = \phi_1^3 \tilde{X}_{t-3} + \phi_1^2 \varepsilon_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t =$$
$$= \ldots = \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}$$

$|\phi_1| < 1 \Rightarrow$ The past '*shocks*' (ε_{t-1}) less and less influence the current observation $\tilde{X}_t$ as the age of the shock (i) increases

$\phi_1 = 1 \Rightarrow$ '*random walk*' – particular type of non-stationary AR(1) process

Moments of an AR(p) process

Mean: For stationary process: $E(X_t) = \mu$

$$\text{Remind}: (1-\sum_{i=1}^{p}\phi_i)\mu = \xi \Rightarrow \mu = \frac{\xi}{(1-\sum_{i=1}^{p}\phi_i)}$$

$$\text{E.g.: stationary AR(1) } (|\phi_1|<1): \mu = \frac{\xi}{(1-\phi_1)}$$

Autocovariance and autocorrelation:

It can be shown (try!) that:

$$\gamma_k = Cov(X_t, X_{t-k}) = E[(X_t-\mu)(X_{t-k}-\mu)] = E\left[\left(\tilde{X}_t\right)\left(\tilde{X}_{t-k}\right)\right]$$
$$= \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \ldots + \phi_p\gamma_{k-p} \qquad k = 1,2,\ldots$$

$$\boxed{\rho_k = \frac{\gamma_k}{\gamma_0} = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \ldots + \phi_p\rho_{k-p} \qquad k = 1,2,\ldots,}$$

The same finite difference equation of the original AR(p) process applies also to the autocovariance and autocorrelation functions

Variance: $\sigma_X^2 = \gamma_0 = \phi_1\gamma_{-1} + \phi_2\gamma_{-2} + \ldots + \phi_p\gamma_{-p} + E\left(\varepsilon_t \widetilde{X}_t\right) =$ (Autocovariance: symmetric function)

$= \phi_1\gamma_1 + \phi_2\gamma_2 + \ldots + \phi_p\gamma_p + \sigma_\varepsilon^2$

Divide the terms by $\sigma_X^2 = \gamma_0$

$$1 = \phi_1\rho_1 + \phi_2\rho_2 + \ldots + \phi_p\rho_p + \frac{\sigma_\varepsilon^2}{\sigma_X^2} \Rightarrow \sigma_X^2 = \frac{\sigma_\varepsilon^2}{1 - \sum_{i=1}^p \phi_i\rho_i}$$

E.g.: stationary AR(1) ($|\phi_1|<1$):

$$\gamma_0 = \sigma_X^2 = \frac{\sigma_\varepsilon^2}{1 - \phi_1\rho_1}$$

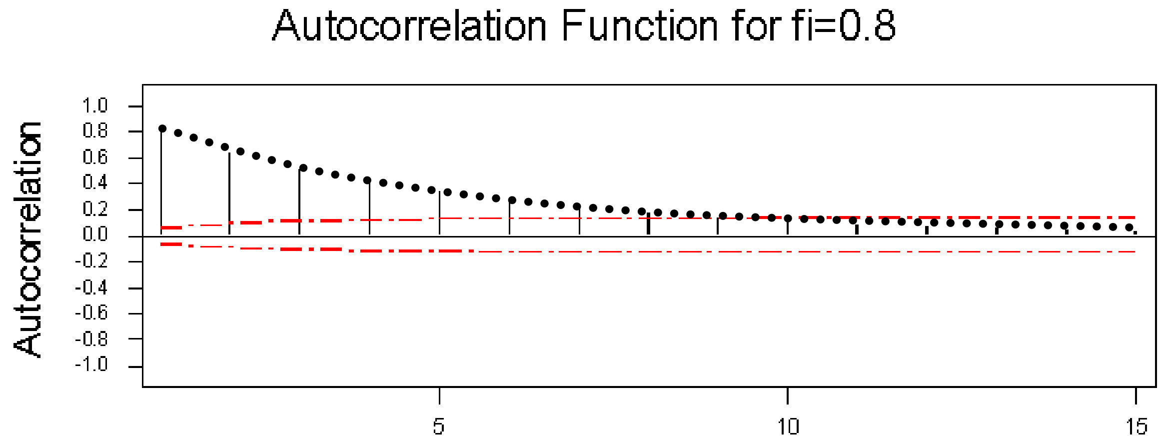
$$\gamma_k = \phi_1\gamma_{k-1} \quad \text{k} = 1,2,\ldots \quad \gamma_1 = \phi_1\gamma_0 = \frac{\phi_1\sigma_\varepsilon^2}{1-\phi_1\rho_1} \quad \gamma_2 = \phi_1\gamma_1 = \frac{\phi_1^2\sigma_\varepsilon^2}{1-\phi_1\rho_1} \quad \ldots \quad \gamma_k = \frac{\phi_1^k\sigma_\varepsilon^2}{1-\phi_1\rho_1}$$

$$\boxed{\rho_k = \frac{\gamma_k}{\gamma_0} \quad \text{k} = 1,2,\ldots,} \qquad \rho_1 = \phi_1 \qquad \rho_2 = \phi_1^2 \qquad \ldots \quad \rho_k = \phi_1^k$$

For a stationary AR(1) process, the autocorrelation function (ACF) "geometrically decays" (if time was 'continuous', the decay would be exponential) $\rho_k = \phi_1^k$

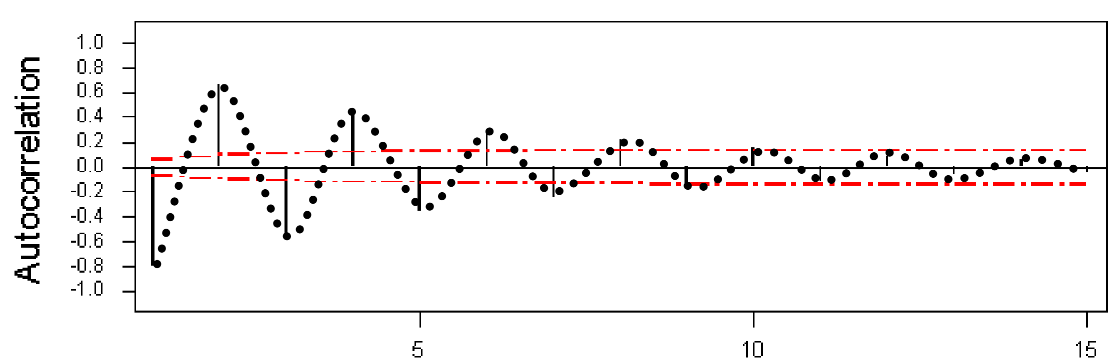
$SACF : \hat{\rho}_k = r_k$

$(\phi_1 = 0.8)$



$SACF : \hat{\rho}_k = r_k$

$(\phi_1 = -0.8)$



Note: It's a general result, true for every stationary *AR(p)* process

How to identify the order p of an $AR(p)$ process?

→ Partial AutoCorrelation Function

$$\tilde{X}_t = \phi_1(\tilde{X}_{t-1}) + \varepsilon_t = \phi_1^2 \tilde{X}_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t = \phi_1^3 \tilde{X}_{t-3} + \phi_1^2 \varepsilon_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t$$

$$\begin{tikzcd} \tilde{X}_t \arrow[r, bend right, "\phi_1"'] \arrow[rr, bend right=70, "\phi_1^2"'] & \tilde{X}_{t-1} \arrow[r, bend right, "\phi_1"'] & \tilde{X}_{t-2} \arrow[r, bend right, "\phi_1"'] & \tilde{X}_{t-3} \end{tikzcd}$$

If ones tries to estimate the coefficients ϕ_{1i} of the AR(1) model in an incremental way:

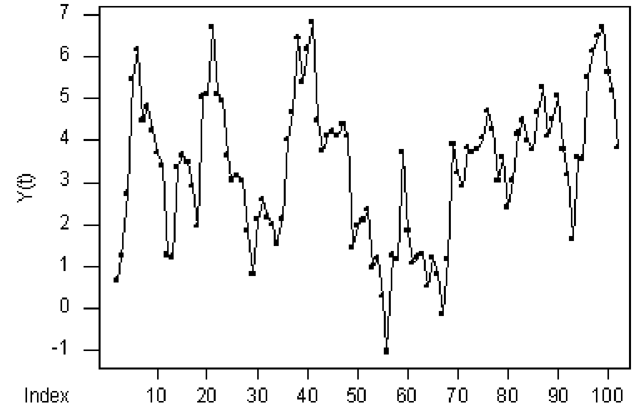
$$\tilde{X}_t = \phi_1(\tilde{X}_{t-1}) + \varepsilon_t$$

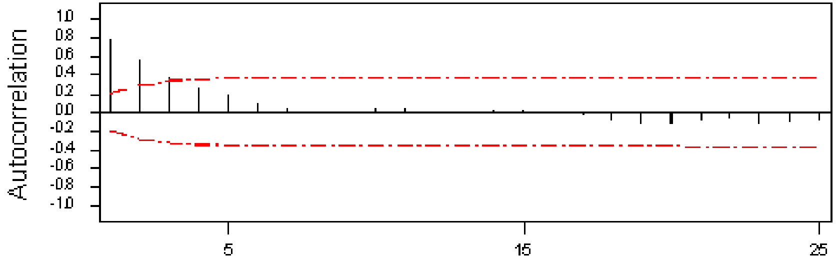
$$\tilde{X}_t = \phi_1^2(\tilde{X}_{t-2}) + \phi_1 \varepsilon_{t-1} + \varepsilon_t$$

$$\tilde{X}_t = \phi_{11}\tilde{X}_{t-1} + \underset{\substack{\downarrow \\ 0}}{\phi_{12}}\tilde{X}_{t-2} + \varepsilon_t$$

Example: AR(1)

$$X_t = \xi + \phi_1 X_{t-1} + \varepsilon_t \quad \xi = 0.5 \;\; \phi_1 = 0.8$$





Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	0.77	7.76	62.08	3	0.37	2.23	109.47	5	0.18	1.02	120.29	7	0.04	0.22	121.48
2	0.56	3.79	94.78	4	0.26	1.48	116.71	6	0.10	0.53	121.30				

Regression Analysis: Y(t) versus Y(t)_1

```
The regression equation is Y(t) = 0.792 + 0.773 Y(t)_1
Predictor          Coef    SE Coef         T        P
Constant         0.7917     0.2337      3.39    0.001
Y(t)_1          0.77308    0.06212     12.44    0.000
S = 1.064         R-Sq = 61.2%     R-Sq(adj) = 60.9%
```

Regression Analysis: Y(t) versus Y(t)_2

The regression equation is: Y(t) = 1.52 + 0.565 Y(t)_2

Predictor	Coef	SE Coef	T	P
Constant	1.5199	0.3070	4.95	0.000
Y(t)_2	0.56527	0.08198	6.90	0.000

S = 1.396 R-Sq = 32.9% R-Sq(adj) = 32.2%

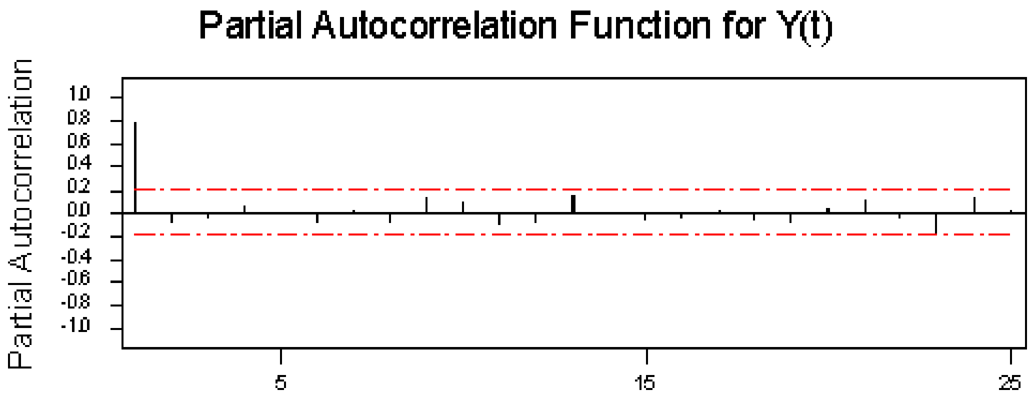
Regression Analysis: Y(t) versus Y(t)_1, Y(t)_2

The regression equation is

Y(t) = 0.858 + 0.848 Y(t)_1 - 0.096 Y(t)_2

Predictor	Coef	SE Coef	T	P
Constant	0.8578	0.2485	3.45	0.001
Y(t)_1	0.8482	0.1020	8.31	0.000
Y(t)_2	-0.0958	0.1014	-0.94	0.347

S = 1.070 R-Sq = 61.0% R-Sq(adj) = 60.2%



Lag	PAC	T	Lag	PAC	T	Lag	PAC	T	Lag	PAC	T
1	0.77	7.76	2	-0.10	-0.97	3	-0.07	-0.66	4	0.05	0.52

$\cong$

Regression Analysis: Y(t) versus Y(t)_1, Y(t)_2

The regression equation is

Y(t) = 0.858 + 0.848 Y(t)_1 - 0.096 Y(t)_2

IDENTIF
↓
AR(1)
↓
FIT
↓
CHECK $e_t \sim N$

Non stationary AR(1) process: random walk wit

Non stationary AR(1) process: random walk with drift

$$X_t = \xi + X_{t-1} + \varepsilon_t$$

$$\xi \neq 0 : "drift"$$

$$X_t = \xi + X_{t-1} + \varepsilon_t = \xi + \xi + X_{t-2} + \varepsilon_{t-1} + \varepsilon_t =$$

$$= 3\xi + X_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t = \ldots = t\xi + \sum_{i=0}^{t-1} B^i \varepsilon_t$$

$$E(X_t) = \mu_t = t\xi$$

$$V(X_t) = t\sigma_\varepsilon^2$$

It's possible to demonstrate that, being $X_0=0$: $\gamma_{k,t} = t\sigma_\varepsilon^2$ and $\rho_k = 1 \quad \forall k$

$$X_t = \xi + X_{t-1} + \varepsilon_t$$

$$\xi \neq 0 : "drift"$$

$$X_t = \xi + X_{t-1} + \varepsilon_t = \xi + \xi + X_{t-2} + \varepsilon_{t-1} + \varepsilon_t =$$

$$= 3\xi + X_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t = \ldots = t\xi + \sum_{i=0}^{t-1} B^i \varepsilon_t$$

$$E(X_t) = \mu_t = t\xi$$

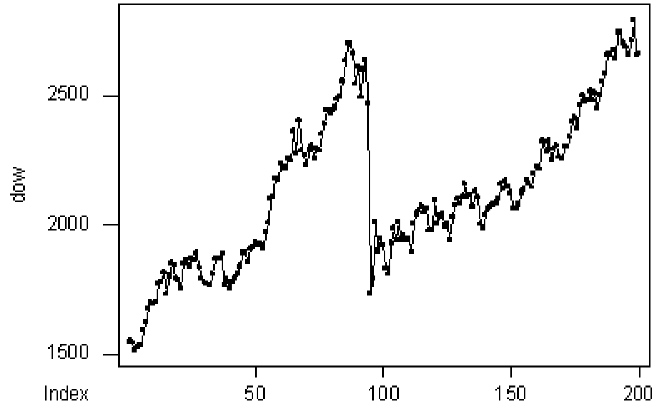
$$V(X_t) = t\sigma_\varepsilon^2$$

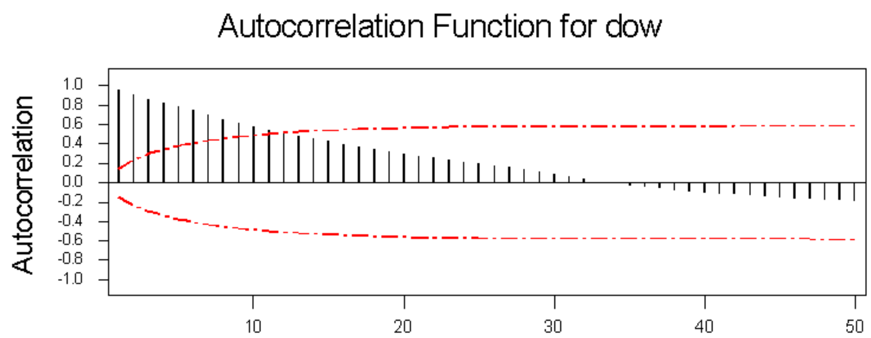
It's possible to demonstrate that, being $X_0=0$: $\gamma_{k,t} = t\sigma_\varepsilon^2$ and $\rho_k = 1 \quad \forall k$

Thus, the ACF does not exhibit a decreasing trend.
Generally speaking, if the ACF (SACF) does not exhibit decreasing trend (or a very slowly decreasing trend) the process is not stationary

Example: Down Jones

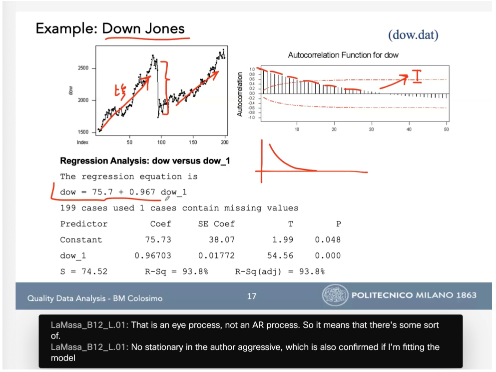
(dow.dat)

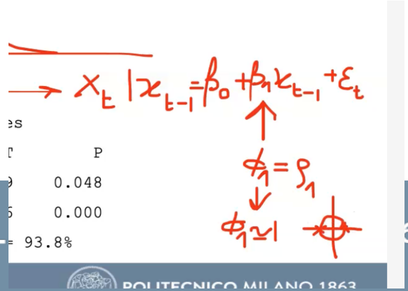




Regression Analysis: dow versus dow_1

```
The regression equation is
dow = 75.7 + 0.967 dow_1
199 cases used 1 cases contain missing values
Predictor          Coef      SE Coef          T          P
Constant          75.73        38.07       1.99      0.048
dow_1           0.96703      0.01772      54.56      0.000
S = 74.52         R-Sq = 93.8%      R-Sq(adj) = 93.8%
```





AR(2):

$$X_t = \xi + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$$

Stationarity conditions (from Jury's test):

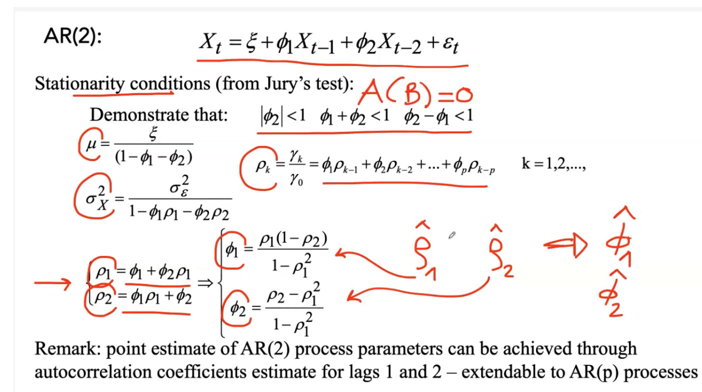
Demonstrate that: $|\phi_2| < 1 \quad \phi_1 + \phi_2 < 1 \quad \phi_2 - \phi_1 < 1$

$$\mu = \frac{\xi}{(1-\phi_1-\phi_2)}$$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \ldots + \phi_p\rho_{k-p} \qquad \text{k} = 1,2,\ldots,$$

$$\sigma_X^2 = \frac{\sigma_\varepsilon^2}{1-\phi_1\rho_1-\phi_2\rho_2}$$

$$\begin{cases} \rho_1 = \phi_1 + \phi_2\rho_1 \\ \rho_2 = \phi_1\rho_1 + \phi_2 \end{cases} \Rightarrow \begin{cases} \phi_1 = \dfrac{\rho_1(1-\rho_2)}{1-\rho_1^2} \\ \\ \phi_2 = \dfrac{\rho_2-\rho_1^2}{1-\rho_1^2} \end{cases}$$



Remark: point estimate of AR(2) process parameters can be achieved through autocorrelation coefficients estimate for lags 1 and 2 – extendable to AR(p) processes

Moving Average models: MA(q)

When the process can be represented by a weighted average of random shocks, ε_t, the process is referred to as a *moving average* process

$$MA(q): X_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} + \varepsilon_t$$

$$\widetilde{X}_t = X_t - \mu = -\theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} + \varepsilon_t$$

"moving average": observation at time t can be estimate as average of random shock at time t (ε_t) and its previous q shocks
(even though the term is improper, as the sum of weights θ_i is not necessarily equal to 1)

With the backshift operator: $\widetilde{X}_t = \left(1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q\right)\varepsilon_t \equiv C(B)\varepsilon_t$

MA(q) processes are always stationary

Moments of a MA(q) process

Mean: $$E(X_t) = E(\mu - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \ldots - \theta_q\varepsilon_{t-q} + \varepsilon_t) = \mu$$

Variance: $$\gamma_0 = \sigma_X^2 = V(X_t) = V(\mu - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \ldots - \theta_q\varepsilon_{t-q} + \varepsilon_t)$$
$$= (1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2)\sigma_\varepsilon^2$$

Autocovariance:

$$\gamma_k = Cov(X_t, X_{t-k}) = E[(X_t - \mu)(X_{t-k} - \mu)] = E\left(\tilde{X}_t\tilde{X}_{t-k}\right) =$$
$$= E\left[\left(\varepsilon_t - \theta_1\varepsilon_{t-1} - \ldots - \theta_q\varepsilon_{t-q}\right)\left(\varepsilon_{t-k} - \theta_1\varepsilon_{t-k-1} - \ldots - \theta_q\varepsilon_{t-k-q}\right)\right]$$

$$\gamma_k = \begin{cases} (-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \ldots + \theta_{q-k}\theta_q)\sigma_\varepsilon^2 & k = 1,2,\ldots,q \\ 0 & k > q \end{cases}$$

Autocorrelation:

$$\boxed{\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} \dfrac{(-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \ldots + \theta_{q-k}\theta_q)}{1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2} & k = 1,2,\ldots,q \\ 0 & k > q \end{cases}}$$

By means of the SACF one can identify a MA process and its order q, depending on the number of lags such that $r_k \neq 0$

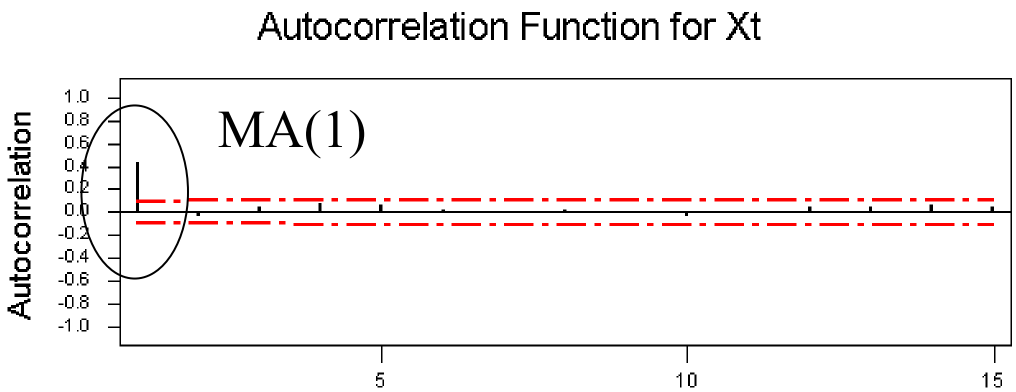
Special cases:

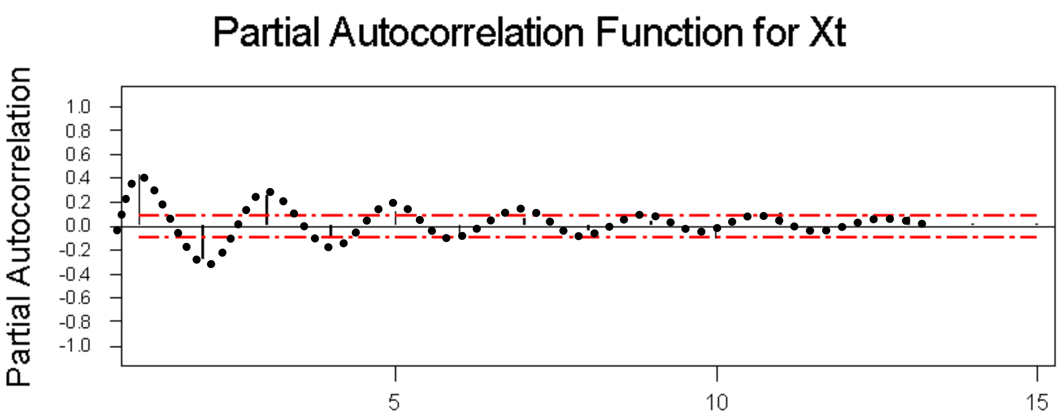
MA(1): $X_t = \mu - \theta_1\varepsilon_{t-1} + \varepsilon_t \quad E(X_t) = \mu \quad \gamma_0 = (1+\theta_1^2)\sigma_\varepsilon^2 \quad \rho_k = \begin{cases} \dfrac{-\theta_1}{1+\theta_1^2} & k = 1 \\ 0 & k > 1 \end{cases}$

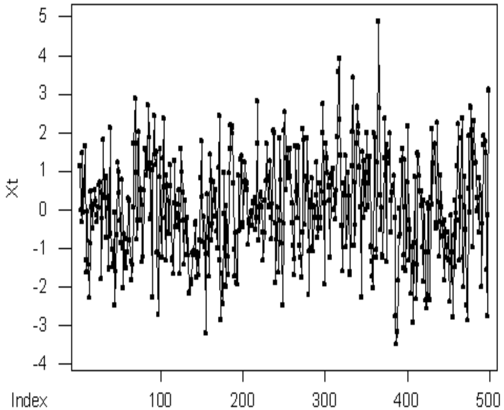
MA(2): $X_t = \mu - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} + \varepsilon_t$

$$E(X_t) = \mu \quad \gamma_0 = (1+\theta_1^2+\theta_2^2)\sigma_\varepsilon^2 \quad \rho_k = \begin{cases} \dfrac{-\theta_1 + \theta_1\theta_2}{1+\theta_1^2+\theta_2^2} & k = 1 \\ \dfrac{-\theta_2}{1+\theta_1^2+\theta_2^2} & k = 2 \\ 0 & k > 2 \end{cases}$$

Initial example: it was a MA(1) process







Example: 500 consecutive observations from a productive process

ARMA(p,q) models

A model that includes both the AR(p) and the MA(q) terms:

$$X_t = \xi + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} + \varepsilon_t$$

$$\tilde{X}_t = X_t - \mu = \phi_1 \tilde{X}_{t-1} + \ldots + \phi_p \tilde{X}_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} + \varepsilon_t$$

$$(1 - \phi_1 B - \ldots - \phi_p B^p)\tilde{X}_t = (1 - \theta_1 B - \ldots - \theta_q B^q)\varepsilon_t$$

$$\boxed{A(B)\widetilde{X}_t = C(B)\varepsilon_t}$$

ARMA process is stationary if its AR part is stationary (it is invertible if its MA part is invertible)

Moments of an ARMA(p,q) process

Mean: $E(X_t) = \mu = \dfrac{\xi}{1 - \sum_{i=1}^{p} \phi_i}$

Autocovariance and autocorrelations (obtained by multiplying by $\tilde{X}_{t-k}$):

$$\boxed{\begin{array}{ll} \gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \ldots + \phi_p \gamma_{k-p} & \mathrm{k} \geq \mathrm{q}+1 \\ \rho_k = \dfrac{\gamma_k}{\gamma_0} = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \ldots + \phi_p \rho_{k-p} & \mathrm{k} \geq \mathrm{q}+1 \end{array}}$$

Apart from the first q lags, the ACF "resembles" the ACF of an AR(p) process.

Generally speaking, one tries to fit an ARMA(p,q) model after trying to fit (unsuccessfully – residual diagnostics) either an AR and/or an MA 'pure' model

ARIMA (p,d,q) models

Most industrial processes are not stationary: when no control action is applied, the process mean tends to departe from the target

A non-stationary ARIMA model exhibit a stationary/non stationary behaviour that depends on the AR term of the model (roots of A(B) polynomial). Particularly:

1. Roots lie strictly outside the unit circle in complex plane: ***stationarity***
2. Roots lie strictly inside the unit circle in complex plane: ***'explosive' non stationarity***
3. Roots lie on the unit circle in complex plane: ***'homogeneous' non stationarity***

Case 2 is pretty rare in production processes.

$$A(B)\widetilde{X}_t = A_p'(B)(1-B)^d\,\widetilde{X}_t = C_q(B)\varepsilon_t$$

Where:

- $A_p'(B)$ is the degree p polynomial (AR term) with all roots falling outside the unit circle;
- d is the degree of the I term (integrated) within the ARIMA model
- $C_q(B)$ is the degree q polynomial of MA term

In order to deal with ARIMA (p,d,q) processes, one has to transform the process into a stationary one by applying the *difference* operator (nabla)

$$\nabla X_t \equiv X_t - X_{t-1} = (1-B)X_t$$

By inverting the ∇ operator one gets the sum operator:

$$SX_t \equiv \sum_{i=-\infty}^{t} X_i = (1+B+B^2+B^3+\ldots)X_t = \frac{X_t}{1-B} = \frac{1}{\nabla}X_t = \nabla^{-1}X_t$$

Under the assumption of infinite number of elements. In the ‘continuous’ domain, integrals replaces sums: this is where the name ‘Integrated’ in ARIMA model originates from

Example 1: random walk

$$X_t = X_{t-1} + \varepsilon_t$$

$$\nabla X_t = X_t - X_{t-1} = \varepsilon_t$$

The process becomes a white noise (stationary)

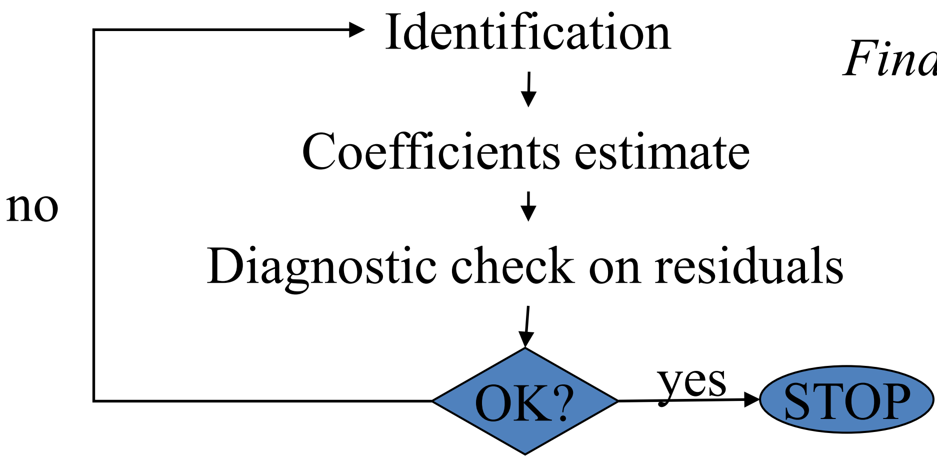
Example 2: ARIMA(0,1,1)=IMA(1,1)

A process that often occurs in industial applications is IMA(1,1)

$$(1-B)X_t = -\theta_1\varepsilon_{t-1} + \varepsilon_t$$

Development of an *ARIMA(p,d,q)* model

Iterative procedure:



Finding the values of p,d,q

$(\hat{\phi}_1,\ldots,\hat{\phi}_p,\hat{\theta}_1,\ldots,\hat{\theta}_q)$

Identification of an *ARIMA(p,d,q)* process

Goal: finding values of p,d,q from a given time series $\{x_t\}$

One has to use the SACF (Sample Auto Correlation Function) and the SPACF (Sample Partial Autocorrelation Function)

Identification: *d parameter*

1. Non-stationarity (*I term*): inferred when the SACF does not exhibit an exponential decay (e.g., linear decay)
2. If the process seems to be non-stationary: apply difference operator. If the resulting time series still exhibit a non-stationary behaviour, iterative application of difference operator is required

Pay attention to "overdifferencing":
Simple way to detect "overdifferencing" consists of computing:

$$Var(x_t), Var(\nabla x_t), \ldots, Var(\nabla^n x_t)$$

Choose *d* such that the variance of the series $\nabla^d x_t$ is minimized

Identification of other parameters (p,q)

Check the SACF and SPACF patterns, reminding that:

- AR(p): the SACF shows "exponential decay" whereas the SPACF is used to choose the degree p
- MA(q): the SACF is used to choose the degree q whereas the SPACF shows "exponential decay"
- ARMA(p,q) resembles an AR(p) after q lags.

Attention: fundamental principle (Box Jenkins): "PARSIMONY"

Most processes can be represented in terms of time series having low values of p and q – they are easier to interpret

Values of p or q equal to (or larger than) 3 are rarely observed in practise

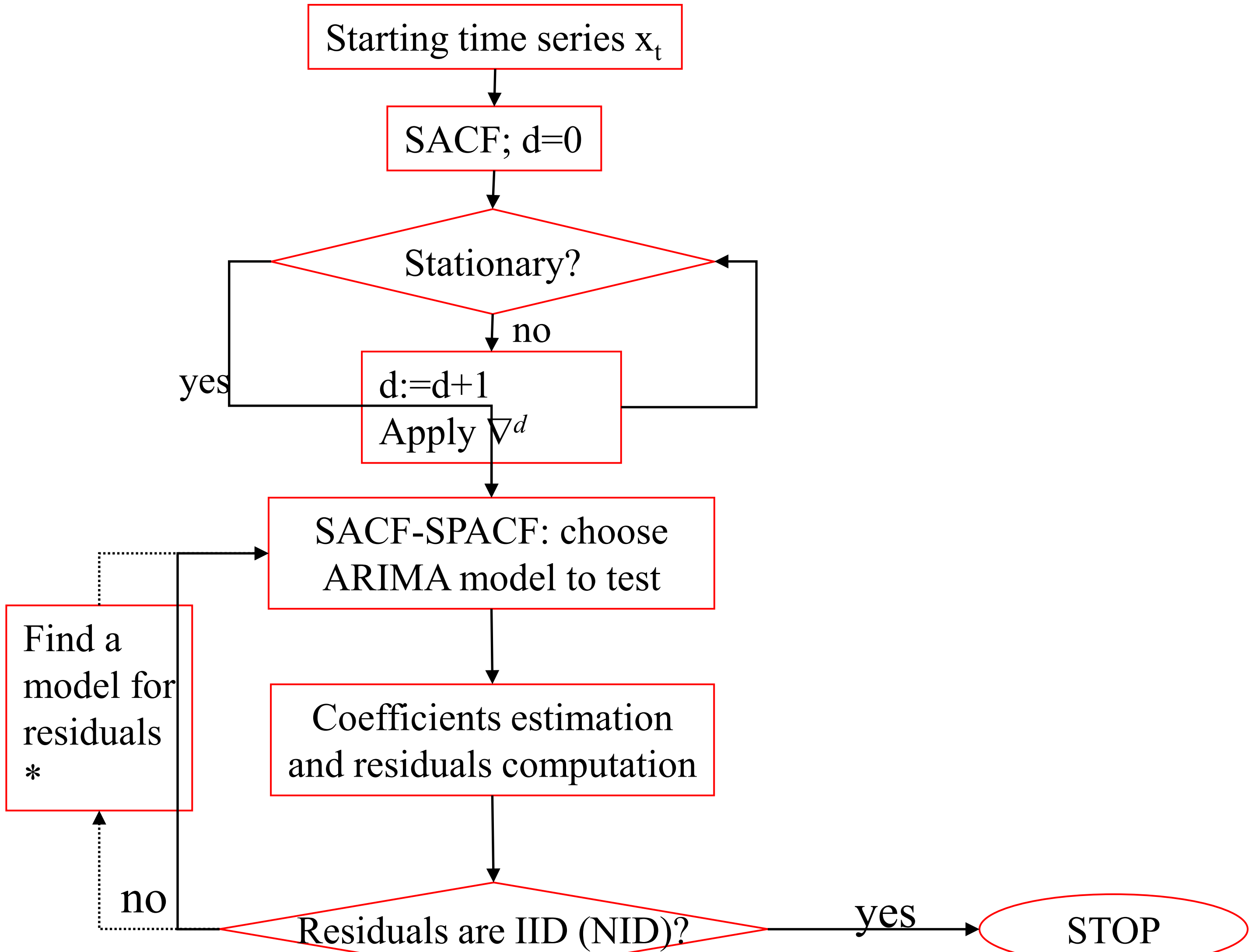
After the identification step (p,d,q) – coefficients estimation

Maximum likelihood method (not covered in this course) –implemented in Minitab

Note:
We noticed that regression (least squares method) can be used for AR(p) models too:
When n is large enough, the two methods converge to the same results

After coefficients estimation – Residuals computation and diagnostics

1. LBQ test
2. Normality test (if required for residual chart – SCC – presented in next slides)


Starting time series x_t
SACF; d=0
Stationary?
no
yes
d:=d+1
Apply ∇^d
SACF-SPACF: choose ARIMA model to test
Find a model for residuals *
Coefficients estimation and residuals computation
no
Residuals are IID (NID)?
yes
STOP

* One can try to identify and fit an ARMA(p_e,q_e) model on residuals

$$\hat{A}_{p_e}(B)e_t = \hat{C}_{q_e}(B)a_t$$

If the model is correct (a_t= IID), such a model can be combined with the original ARIMA(p,d,q) model, as follows:

$$\hat{A}(B)\nabla^d \tilde{X}_t = \hat{C}(B)e_t \qquad e_t = \frac{\hat{C}_{q_e}(B)}{\hat{A}_{p_e}(B)}a_t$$

$$\hat{A}(B)\nabla^d \tilde{X}_t = \hat{C}(B)\frac{\hat{C}_{q_e}(B)}{\hat{A}_{p_e}(B)}a_t$$

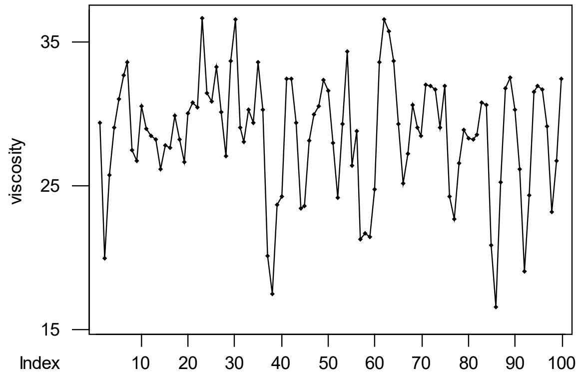
$$\hat{A}(B)\hat{A}_{p_e}(B)\nabla^d \tilde{X}_t = \hat{C}(B)\hat{C}_{q_e}(B)a_t$$

The result is an ARIMA($p+p_e$,d,$q+q_e$) – pay attention to the parsimony principle

Example: viscosity data

Viscosity of a chemical product

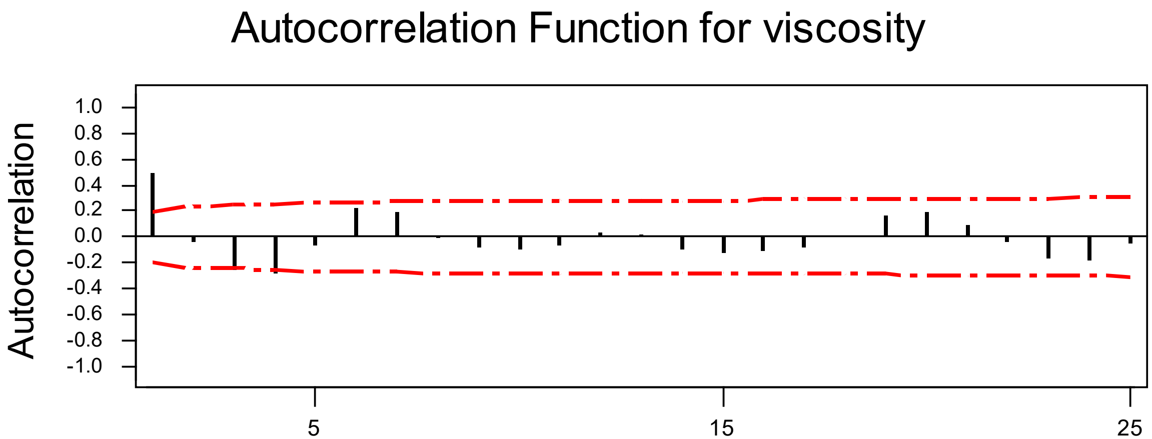
(source: Montgomery D.C., Johnson, L.A., Gardiner, J.S., Forecasting and Time Series Analysis – McGraw-Hill)



Runs Test: viscosity

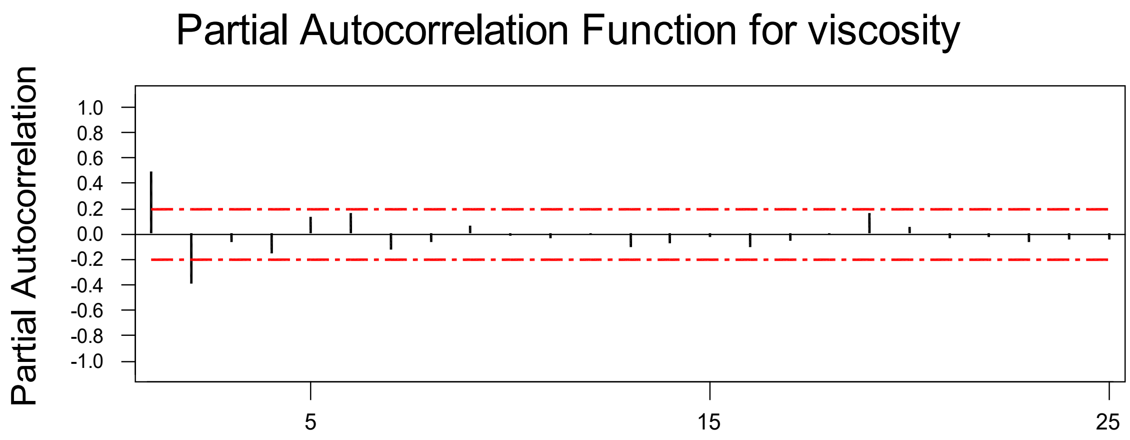
Stationarity seems ok

```
viscosit
K =    28.5696
The observed number of runs =  37
The expected number of runs =  50.0200
57 Observations above K   43 below
          The test is significant at  0.0076
```



Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	0.49	4.94	25.13	8	-0.01	-0.08	51.19	15	-0.12	-0.87	57.41	22	-0.04	-0.30	68.37
2	-0.05	-0.41	25.39	9	-0.09	-0.66	52.16	16	-0.11	-0.78	58.95	23	-0.17	-1.12	72.13
3	-0.26	-2.16	32.73	10	-0.11	-0.75	53.44	17	-0.09	-0.62	59.96	24	-0.19	-1.22	76.74
4	-0.28	-2.22	41.25	11	-0.07	-0.53	54.08	18	0.00	0.03	59.96	25	-0.06	-0.41	77.27
5	-0.07	-0.54	41.82	12	0.03	0.22	54.19	19	0.16	1.08	63.06				
6	0.22	1.63	46.99	13	0.01	0.09	54.21	20	0.18	1.24	67.29				
7	0.20	1.42	51.18	14	-0.11	-0.74	55.54	21	0.08	0.53	68.12				

- Stationarity seems ok
- SACF seems to exhibit an exponential decay



Lag	PAC	T	Lag	PAC	T	Lag	PAC	T	Lag	PAC	T
1	0.49	4.94	8	-0.06	-0.62	15	-0.02	-0.22	22	-0.03	-0.27
2	-0.39	-3.88	9	0.06	0.64	16	-0.10	-1.00	23	-0.07	-0.65
3	-0.06	-0.60	10	-0.02	-0.18	17	-0.05	-0.52	24	-0.04	-0.42
4	-0.15	-1.50	11	-0.03	-0.34	18	-0.00	-0.05	25	-0.05	-0.46
5	0.13	1.33	12	-0.00	-0.03	19	0.16	1.64			
6	0.16	1.62	13	-0.11	-1.05	20	0.05	0.53			
7	-0.13	-1.26	14	-0.07	-0.71	21	-0.04	-0.36			

Let's try with an AR(2)

ARIMA Model: viscosity

ARIMA model for viscosity

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	0.7187	0.0923	7.79	0.000
AR 2	-0.4344	0.0922	-4.71	0.000
Constant	20.5061	0.3279	62.53	0.000
Mean	28.6516	0.4582		

Number of observations: 100

Residuals: SS = 1042.78 (backforecasts excluded)

MS = 10.75 DF = 97

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	12.2	21.1	28.3	38.0
DF	9	21	33	45
P-Value	0.202	0.451	0.701	0.759

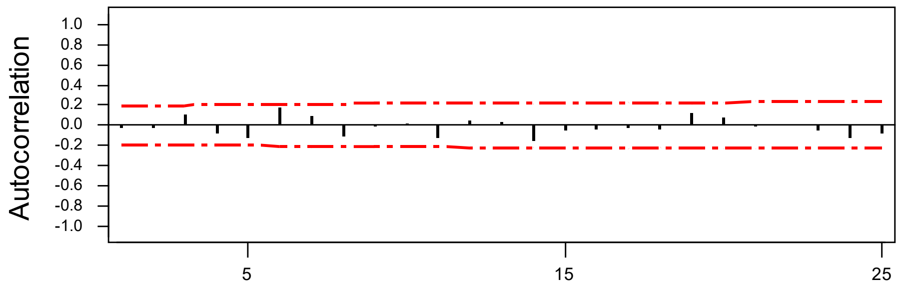
Estimated model $\boxed{X_t = 20.506 + 0.7187 X_{t-1} - 0.4344 X_{t-2}}$

Estimated model $\boxed{X_t = 20.506 + 0.7187 X_{t-1} - 0.4344 X_{t-2}}$

Regression Analysis: viscosity versus viscosity_1, viscosity_2

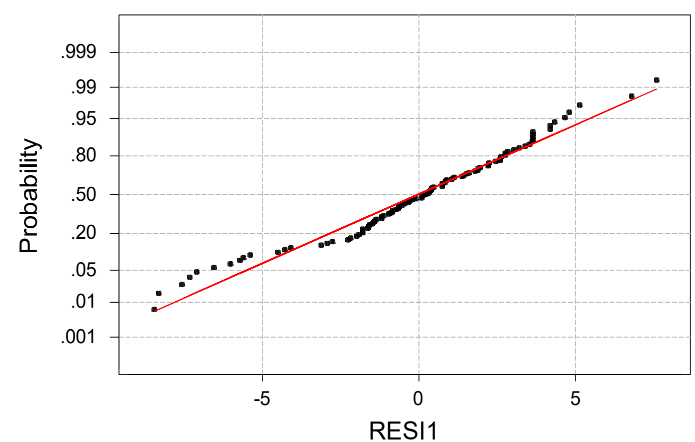
```
The regression equation is
viscosity = 20.1 + 0.707 viscosity_1 - 0.406 viscosity_2
98 cases used 2 cases contain missing values

Predictor        Coef    SE Coef       T       P
Constant       20.081      2.613    7.68   0.000
Viscos_t-1    0.70672    0.09112    7.76   0.000
Viscos_t-2   -0.40594    0.09119   -4.45   0.000

S = 3.223      R-Sq = 38.9%     R-Sq(adj) = 37.6%
```

Autocorrelation Function for RESI1



Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	-0.02	-0.25	0.06	8	-0.12	-1.09	9.89	15	-0.05	-0.45	15.67	22	0.00	0.04	18.58
2	-0.02	-0.22	0.11	9	-0.01	-0.11	9.90	16	-0.05	-0.44	15.97	23	-0.06	-0.49	19.00
3	0.11	1.08	1.33	10	0.01	0.10	9.92	17	-0.03	-0.26	16.07	24	-0.13	-1.09	21.13
4	-0.09	-0.91	2.25	11	-0.14	-1.25	12.03	18	-0.04	-0.35	16.27	25	-0.08	-0.71	22.08
5	-0.13	-1.29	4.12	12	0.04	0.34	12.20	19	0.11	0.96	17.76				
6	0.17	1.68	7.43	13	0.03	0.24	12.28	20	0.08	0.69	18.56				
7	0.09	0.86	8.36	14	-0.16	-1.46	15.36	21	-0.01	-0.09	18.58				

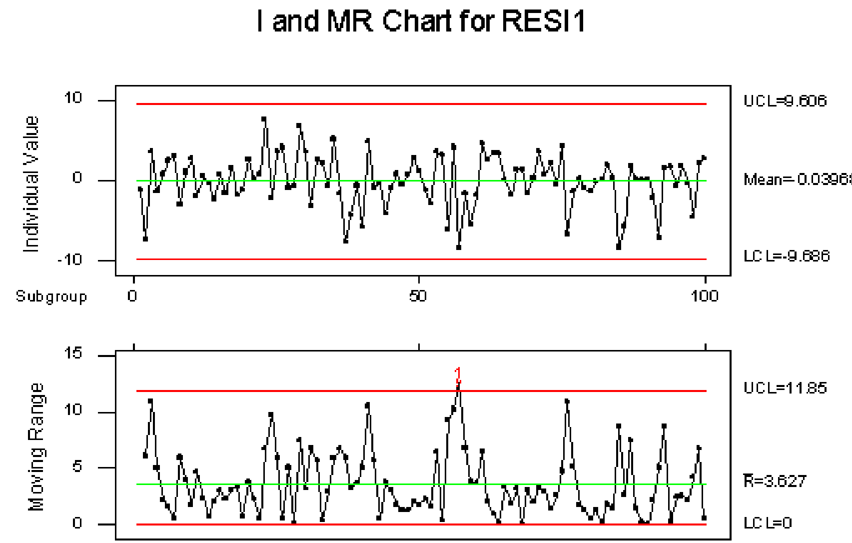
Normal Probability Plot



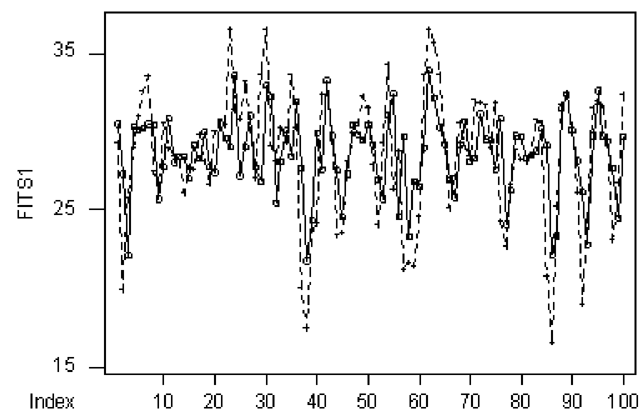
Average: -0.0396777
StDev: 3.24524
N: 100

Anderson-Darling Normality Test
A-Squared: 0.966
P-Value: 0.014

SCC

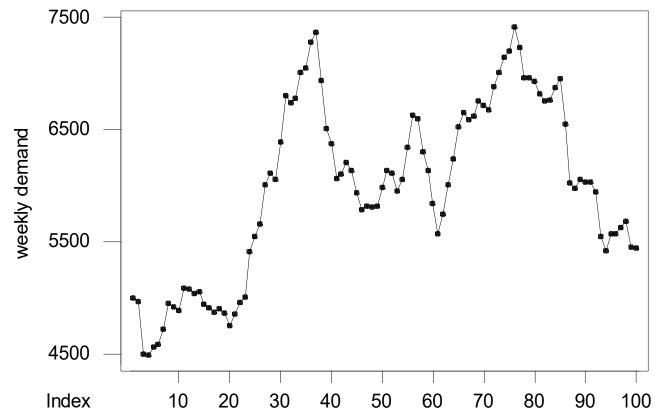

I and MR Chart for RESI1
Individual Value
10
0
-10
UCL=9.606
Mean=0.03968
LCL=-9.686
Subgroup
0
50
100
Moving Range
15
10
5
0
1
UCL=11.85
R̄=3.627
LCL=0

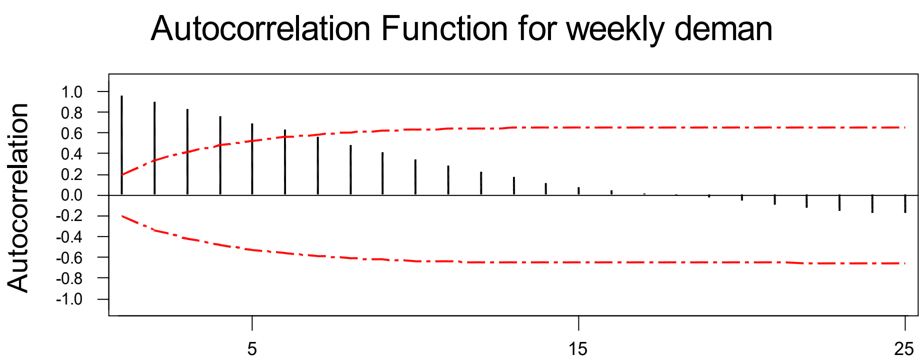
FVC


FITS1
35
25
15
Index
10
20
30
40
50
60
70
80
90
100

100 data – weekly demand of plastic containers from plastic injection moulding (drug production sector)

(source: Montgomery D.C., Johnson, L.A., Gardiner, J.S., Forecasting and and Time Series Analysis – McGraw-Hill)

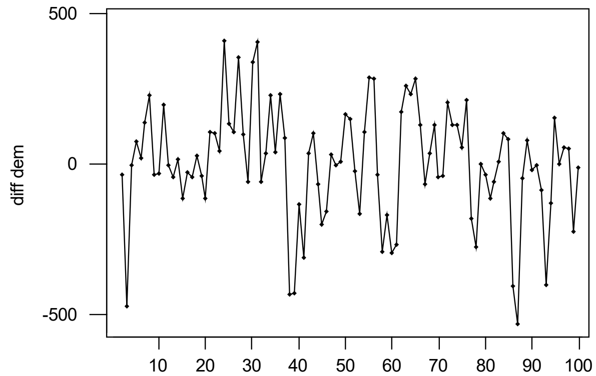




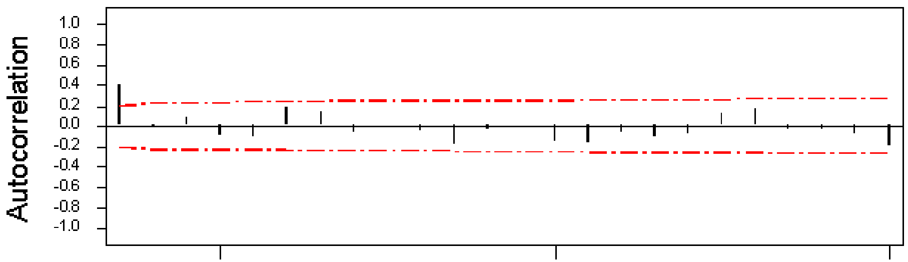
Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	0.96	9.61	95.13	8	0.48	1.58	466.81	15	0.08	0.23	520.48	22	-0.13	-0.38	524.39
2	0.90	5.32	179.07	9	0.41	1.32	486.06	16	0.05	0.14	520.75	23	-0.16	-0.47	527.61
3	0.83	3.93	251.61	10	0.34	1.08	499.54	17	0.02	0.05	520.78	24	-0.17	-0.53	531.68
4	0.76	3.14	312.62	11	0.28	0.87	508.60	18	-0.01	-0.02	520.79	25	-0.18	-0.54	535.97
5	0.69	2.61	363.79	12	0.23	0.70	514.59	19	-0.03	-0.08	520.87				
6	0.63	2.23	406.78	13	0.17	0.53	518.10	20	-0.05	-0.17	521.26				
7	0.56	1.89	440.93	14	0.12	0.36	519.78	21	-0.09	-0.28	522.34				

It looks like a non-stationary process

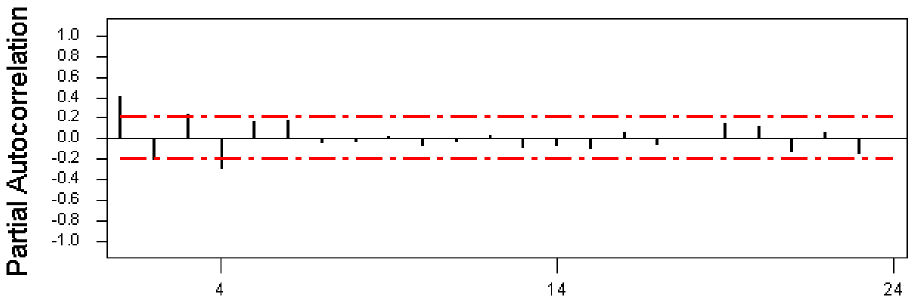
Let’s apply the ∇ operator:



Autocorrelation Function for diff dem



Partial Autocorrelation Function for diff dem



After differencing: MA(1)

Thus, model IMA(1,1):

$$A(B)\widetilde{X}_t = A_p'(B)(1-B)^d \widetilde{X}_t = C_q(B)\varepsilon_t$$

ARIMA Model: weekly demand

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
MA 1	-0.7331	0.0688	-10.66	0.000

Differencing: 1 regular difference

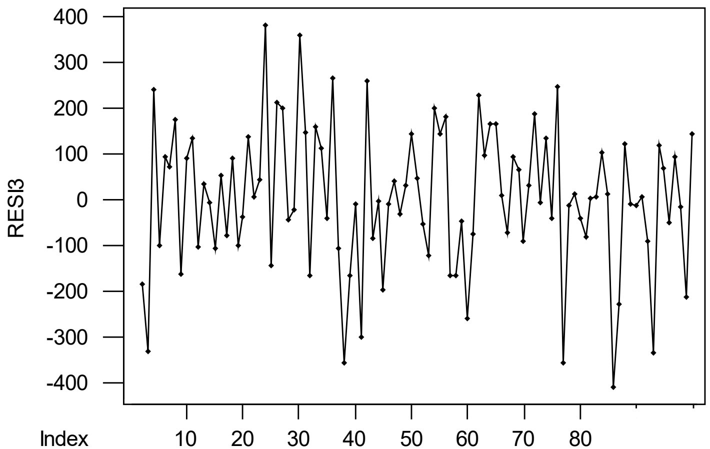
Number of observations: Original series 100, after differencing 99

Residuals: SS = 2405478 (backforecasts excluded)

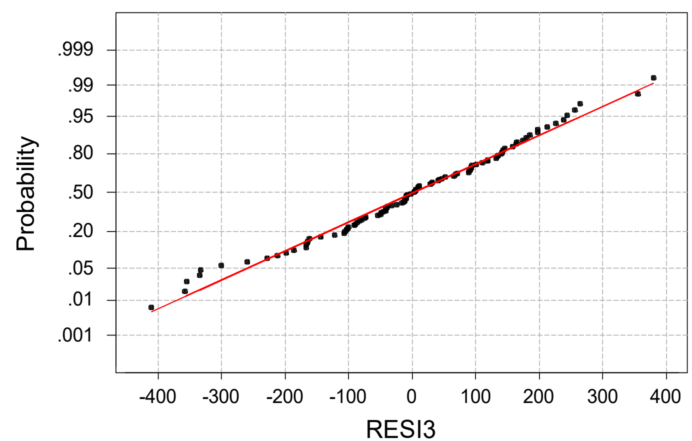
MS = 24546 DF = 98

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	21.6	41.1	67.8	89.7
DF	11	23	35	47
P-Value	0.028	0.012	0.001	0.000



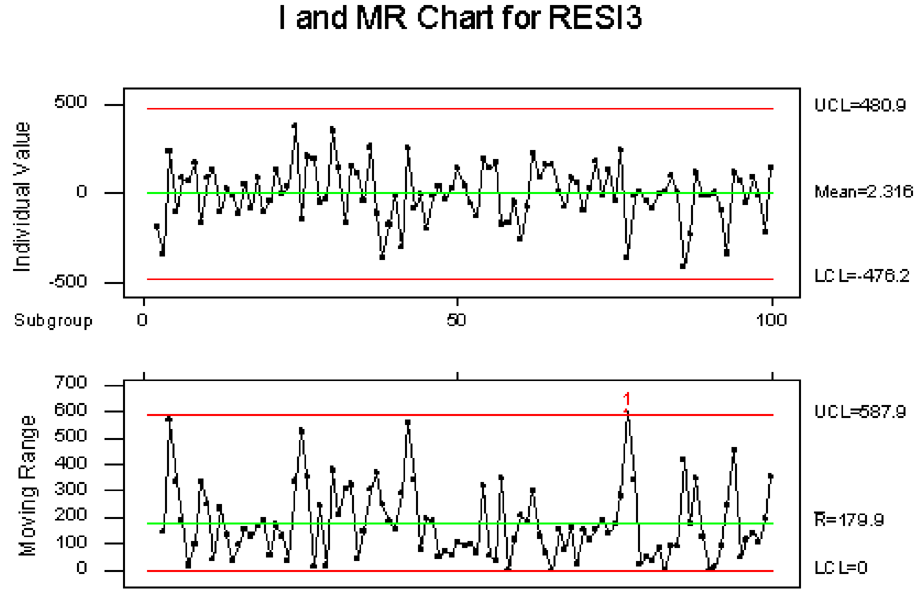
Normal Probability Plot



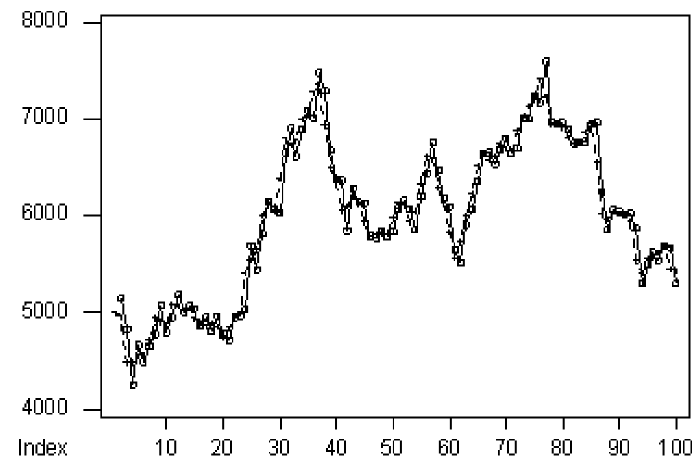
Average: 2.31596
StDev: 156.653
N: 99

Anderson-Darling Normality Test
A-Squared: 0.383
P-Value: 0.390

SCC


I and MR Chart for RESI3
Individual Value
500
0
-500
UCL=480.9
Mean=2.316
LCL=-476.2
Subgroup
0
50
100
Moving Range
700
600
500
400
300
200
100
0
1
UCL=587.9
R=179.9
LCL=0

FVC


8000
7000
6000
5000
4000
Index
10
20
30
40
50
60
70
80
90
100

Additional slides

Moments of an AR(p) process

Mean: For stationary process: $E(X_t) = \mu$

$$\text{Remind}: (1-\sum_{i=1}^{p}\phi_i)\mu = \xi \Rightarrow \mu = \frac{\xi}{(1-\sum_{i=1}^{p}\phi_i)}$$

E.g.: stationary AR(1) ($|\phi_1|<1$): $\mu = \dfrac{\xi}{(1-\phi_1)}$

$$E(\varepsilon_i \varepsilon_j) = \begin{cases} 0 & \text{for } i \neq j \\ \sigma_\varepsilon^2 & \text{for } i = j \end{cases}$$

Autocovariance and autocorrelation:

0

$$\begin{aligned}
\gamma_k &= Cov(X_t, X_{t-k}) = E[(X_t-\mu)(X_{t-k}-\mu)] = E[(\tilde{X}_t)(\tilde{X}_{t-k})] = \\
&= E[(\phi_1 \tilde{X}_{t-1} + \phi_2 \tilde{X}_{t-2} + \ldots + \phi_p \tilde{X}_{t-p} + \varepsilon_t)(\tilde{X}_{t-k})] = \\
&= E(\phi_1 \tilde{X}_{t-1}\tilde{X}_{t-k}) + E(\phi_2 \tilde{X}_{t-2}\tilde{X}_{t-k}) + \ldots + E(\phi_p \tilde{X}_{t-p}\tilde{X}_{t-k}) + E(\varepsilon_t \tilde{X}_{t-k}) = \\
&= \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \ldots + \phi_p\gamma_{k-p} \qquad \text{k} = 1,2,\ldots
\end{aligned}$$

$$\boxed{\rho_k = \frac{\gamma_k}{\gamma_0} = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \ldots + \phi_p\rho_{k-p} \qquad \text{k} = 1,2,\ldots,}$$

Invertibility of a MA(q) process – (briefly)

Note: the *invertibility* concept for MA(q) processes is analogous to the stationarity concept for AR(p) processes

A time series process is invertible if it can be expressed as an AR process as follows:

$$\tilde{X}_t = \pi_1 \tilde{X}_{t-1} + \pi_2 \tilde{X}_{t-2} + \ldots + \varepsilon_t$$

Where the sum is allowed to have an infinite number of terms, but it must converge to a finite value

Example - MA(1): $\tilde{X}_t = -\theta_1 \varepsilon_{t-1} + \varepsilon_t = (1 - \theta_1 B)\varepsilon_t$

$$\varepsilon_t = \frac{1}{1-\theta_1 B}\tilde{X}_t = \sum_{i=0}^{\infty} \theta_1{}^i B^i \tilde{X}_t$$

Invertibility condition for a MA(1) process: $|\theta_1| < 1$

Moments of a MA(q) process

Mean: $$E(X_t) = E(\mu - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \ldots - \theta_q\varepsilon_{t-q} + \varepsilon_t) = \mu$$

Variance: $$\gamma_0 = \sigma_X^2 = V(X_t) = V(\mu - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \ldots - \theta_q\varepsilon_{t-q} + \varepsilon_t) =$$
$$= (1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2)\sigma_\varepsilon^2$$

Autocovariance: $$\gamma_k = Cov(X_t, X_{t-k}) = E[(X_t - \mu)(X_{t-k} - \mu)] = E\left(\tilde{X}_t \tilde{X}_{t-k}\right) =$$
$$= E[(\varepsilon_t - \theta_1\varepsilon_{t-1} - \ldots - \theta_q\varepsilon_{t-q})(\varepsilon_{t-k} - \theta_1\varepsilon_{t-k-1} - \ldots - \theta_q\varepsilon_{t-k-q})]$$

$$E(\varepsilon_i\varepsilon_j) = \begin{matrix} 0 & \text{for } i \neq j \\ \sigma_\varepsilon^2 & \text{for } i = j \end{matrix}$$

for $k < q$

$$= E\Big[\left(\varepsilon_t - \theta_1\varepsilon_{t-1} - \ldots - \theta_k\varepsilon_{t-k} - \theta_{k+1}\varepsilon_{t-(k+1)} - \ldots - \theta_q\varepsilon_{t-q}\right)$$
$$\left(\varepsilon_{t-k} - \theta_1\,\varepsilon_{t-(k+1)} - \ldots - \theta_{q-k}\varepsilon_{t-(k+q-k)} - \ldots - \theta_{q-1}\varepsilon_{t-(k+q-1)} - \theta_q\varepsilon_{t-(k+q)}\right)\Big]$$

$$\gamma_k = \begin{cases} (-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \ldots + \theta_{q-k}\theta_q)\sigma_\varepsilon^2 & k = 1, 2, \ldots, q \\ 0 & k > q \end{cases}$$