La procedura Box-Jenkins

La selezione del modello

- Procedura di Box e Jenkins (1976): procedura per costruire, a partire dall'osservazione dei dati, un modello ARMA atto ad approssimare il processo generatore della serie storica.

- Si ipotizza che la serie sia già stata resa stazionaria attraverso una serie di trasformazioni preliminari (se necessarie)

Identificazione

⇓

Stima dei parametri

⇓

Controllo diagnostico

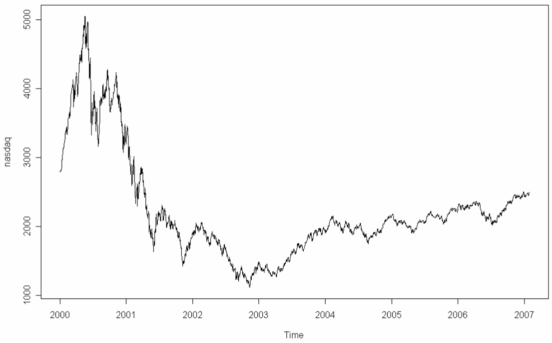
- Quando il controllo diagnostico non è soddisfacente il modello deve essere *rispecificato* per cui le tre fasi principali possono essere ripetute più volte in maniera iterativa;
- Quando il modello risulta soddisfacente si passa alla fase del suo *utilizzo* per scopi descrittivi e previsivi.

Identificazione

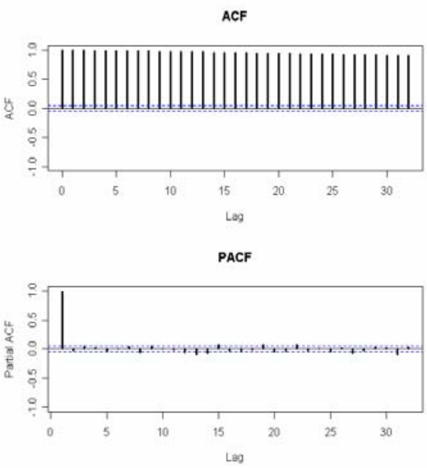
-Specificazione dell’ordine del modello (parametri *p* e *q*)

-Principali strumenti da usare: ACF e PACF stimate

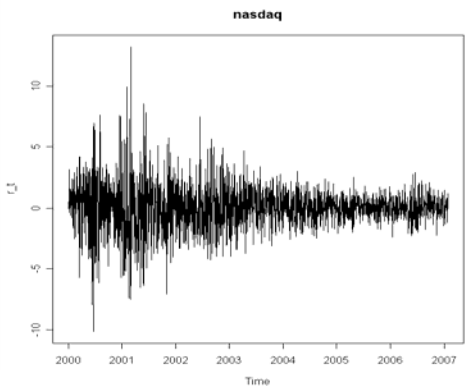
-Se le autocorrelazioni tendono ad annullarsi molto lentamente (o non si annullano affatto) è probabile che il processo generatore sia non stazionario (Es: prezzi degli strumenti finanziari, ved. correlogramma su ENI)



Esempio NASDAQ



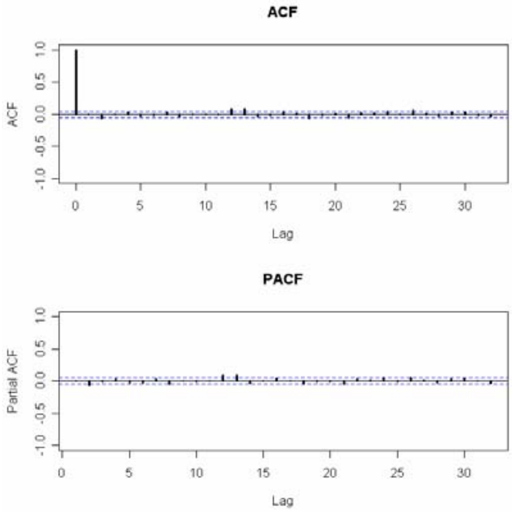
…passando ai rendimenti si ha:



Struttura tipica di un processo non stazionario…

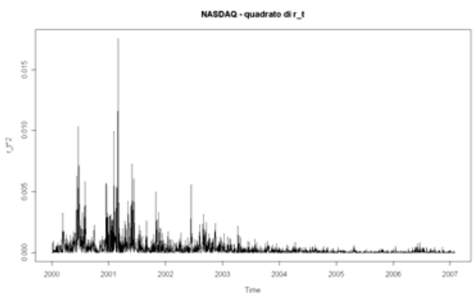
Esempio NASDAQ (2)

… il cui correlogramma è

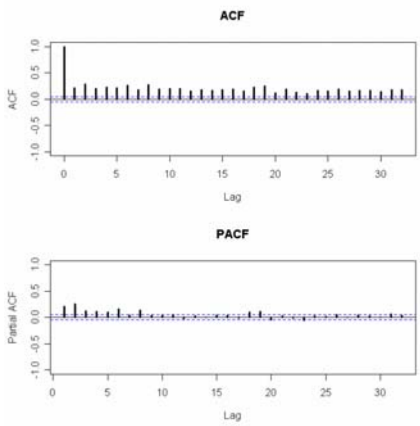


Esempio NASDAQ (3)

Se eleviamo al quadrato i rendimenti abbiamo



…il cui correlogramma è:

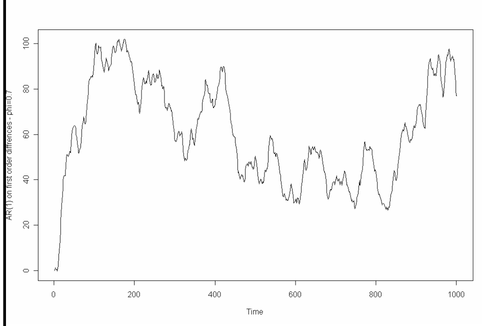


Operazioni preliminari

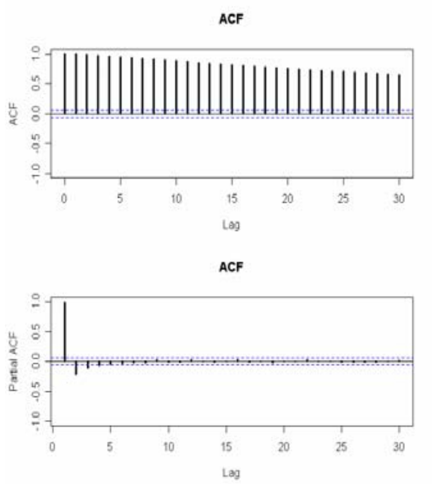
Serie generata da un PS ARIMA(1,1,0), cioè

$$\Delta Y_t = \phi \Delta Y_{t-1} + \varepsilon_t$$

$$y_t - y_{t-1} = 0.7(y_{t-1} - y_{t-2}) + \varepsilon_t$$



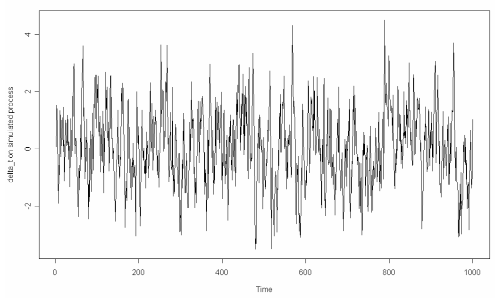
L'impressione è quella di una serie non stazionaria in media.



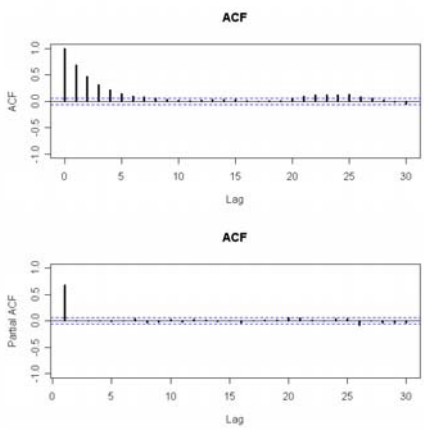
Il correlogramma conferma l'impressione di non stazionarietà avuta dall'osservazione del grafico della serie.

Operazioni preliminari (2)

… differenziamo la serie e otteniamo il seguente grafico



…il cui correlogramma è il seguente



che è il correlogramma tipico di un processo AR(1).

La fase di stima dei parametri

-Minimi quadrati iterati o massima verosimiglianza (ipotesi di normalità)

-Fase complicata numericamente e affidata ai calcolatori

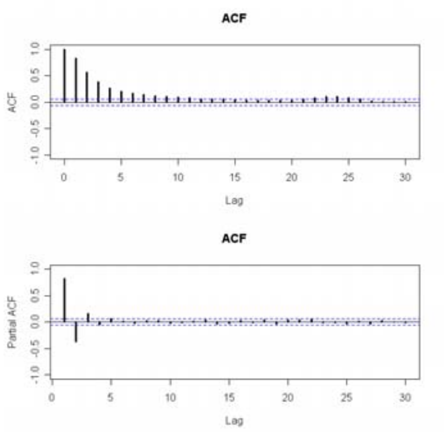
-Le stime dei parametri si indicano con il cappello (*hat*) per cui un modello ARMA(1,1) può essere scritto:

$$\hat{y}_t = \hat{c} + \hat{\phi}_1 y_{t-1} + \hat{\theta}_1 \hat{\varepsilon}_{t-1}$$

dove $\hat{y}_t$ è il valore stimato al tempo *t*.

Stima dei parametri (2)

Correlogramma di una serie generata da un processo ARMA(1,1)



Si può pensare ad un MA(*q*)…

…proviamo un MA(6).

	coef	s.e.	t.stat	p.value
ma1	1.205855	0.031777	37.94776	1.90E-195
ma2	0.852527	0.048814	17.46498	9.05E-60
ma3	0.546125	0.052003	10.50179	1.55E-24
ma4	0.31637	0.050841	6.222763	7.19E-10
ma5	0.134581	0.047594	2.827682	0.004783
ma6	0.037023	0.029439	1.257597	0.208833
intercept	0.047095	0.132908	0.354343	0.723157

La significatività dei parametri

Dopo la fase di stima è necessario verificare se i coefficienti stimati sono significativamente diversi da zero.

$H_0: \phi_i = 0$

$H_1: \phi_i \neq 0$ $\qquad \dfrac{\hat{\phi}_i - 0}{\sqrt{\mathrm{var}(\hat{\phi}_i)}} \sim t_{n-m}$

dove m è il numero di parametri stimati.

***Rule of thumb*:**

$\alpha = 0.05$

$\left| \dfrac{\hat{\phi}_i}{\sqrt{\mathrm{var}(\hat{\phi}_i)}} \right| > 2$ il parametro stimato è significativamente diverso da zero.

Osservazione:

$\sqrt{\mathrm{var}(\hat{\phi}_i)}$ = standard error (s.e.) di $\hat{\phi}_i$

Il criterio della parsimonia

I polinomi AR: $\left(1-\phi_1 L - \ldots - \phi_p L^p\right)$

e MA: $\left(1-\theta_1 L - \ldots - \theta_q L^q\right)$

non devono avere radici comuni.

Esempio. ARMA(1,1)

$$(1-0.7L)Y_t = (1-0.7L)\varepsilon_t$$

$$\phi_1 = 0.7 \ e \ \theta_1 = -0.7$$

$$\Downarrow$$

$Y_t = \varepsilon_t$ la serie è W. N.

In generale, se nel modello ARMA(p,q)

$\phi_k \cong -\theta_k$ k=1,2,…,min(p,q)
allora il modello e **sovraparametrizzato**.

Esempio 1: non significativo il parametro di ordine superiore

$$\Delta Y_t = \left(1 - \underset{(0.1)}{0.86}\, L + \underset{(0.1)}{0.02}\, L^2\right)\varepsilon_t$$

standard errors

$$\left|\frac{-0.86}{0.1}\right| > 2 \Rightarrow \text{significativo}$$

$$\left|\frac{0.02}{0.1}\right| < 2 \quad \Rightarrow \text{non significativo}$$

Il modello finale sarà il seguente:

$$\Delta Y_t = (1 - 0.86B)\varepsilon_t$$

Esempio 2: non significativo il parametro di ordine inferiore

$$\Delta Y_t = \left(1 - \underset{(0.1)}{0.02}\, L - \underset{(0.1)}{0.86}\, L^2\right)\varepsilon_t$$

E' significativo θ_2 , ma non θ_1

Osservazione. Solitamente i parametri di ordine inferiore sono indispensabili nella procedura di stima, di conseguenza occorre cautela prima di eliminarli.

Bisogna analizzare $\rho(\hat{\theta}_1, \hat{\theta}_2)$

Se $\rho(\hat{\theta}_1, \hat{\theta}_2) \cong 1$

anche la loro stima è strettamente correlata, di conseguenza è possibile ottenere buoni risultati (in termini di descrizione dei dati) anche usando un solo parametro anziché due.

Se, ad esempio $\rho(\hat{\theta}_1, \hat{\theta}_2) = 0.94$

il modello diventa:

$$\Delta Y_t = (1 - 0.88L)\varepsilon_t$$

Riprendendo l'esempio su ARMA(1,1) simulato: parsimonia

…avevamo provato un MA(6)

	coef	s.e.	t.stat	p.value
ar1	0.6975	0.0253	27.5878	0.0000
ma1	0.4929	0.0298	16.5447	0.0000
intercept	0.0449	0.1597	0.2812	0.7786

-Modello ARMA(1,1) più flessibile e "parsimonioso"

Esempio su NASDAQ: parsimonia

Proviamo un AR(4) sul quadrato dei rendimenti del NASDAQ

	coef	s.e.	t.stat	p.value
ar1	0.1128	0.023135	4.8762	0.0000
ar2	0.2086	0.023146	9.0104	0.0000
ar3	0.1053	0.023148	4.5483	0.0000
ar4	0.1162	0.023124	5.0230	0.0000
intercept	0.0004	0.000048	7.5611	0.0000

…proviamo quindi un ARMA(1,1):

	coef	s.e.	t.stat	p.value
ar1	0.9837	0.00598	164.39	0.0000
ma1	-0.9022	0.01526	-59.10	0.0000
intercept	0.0004	0.00011	3.23290	0.0012

Anche qui il rischio è quello di sovraparametrizzazione…

Controllo diagnostico sui residui

-Verifica dell'adeguatezza del modello stimato: analisi dei residui

$\hat{\varepsilon}_t = y_t - \hat{y}_t$

-Presenza di autocorrelazione residua. H_0: $\hat{\varepsilon}_t \sim WN$

Test sui singoli valori della funzione di autocorrelazione stimata sui residui, cioè

$$\rho_k(\hat{\varepsilon}) = \frac{\sum_{t=k+1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_{t-k}}{\sum_{t=1}^{T} \hat{\varepsilon}_t^2}$$

-Verifica complessiva di assenza di autocorrelazione nei residui con il test Q di Ljung-Box:

H_0: assenza di correlazione nei residui fino al lag *m*.

H_1: presenza di correlazione nei residui fino al lag *m*.

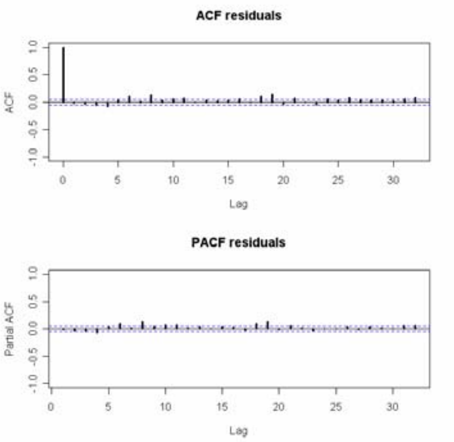
Statistica-test:

$$Q(m) = T \cdot (T+2) \sum_{k=1}^{m} \frac{\rho_k^2(\hat{\varepsilon})}{T-k}$$ Sotto l'ipotesi nulla $Q(m) \sim \chi_m^2$

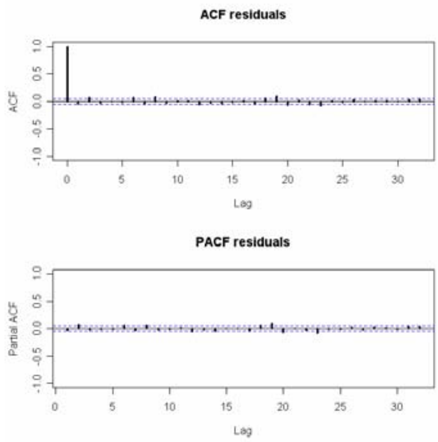
Esempio su NASDAQ (continua)

Correlogramma dei residui con i due modelli stimati sui quadrati dei rendimenti del NASDAQ.

Residui del modello AR(4)

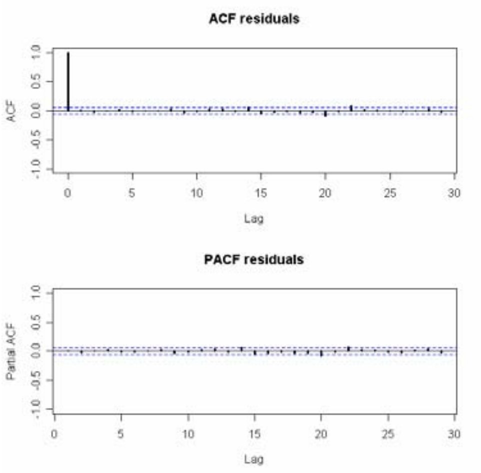


Residui del modello ARMA(1,1)



Esempio su ARMA(1,1) simulato

Correlogramma dei residui con il modello adeguato su una serie simulata da un ARMA(1,1)



Controllo sull'omoschedasticità dei residui

Costanza della varianza dei residui nel tempo.
In caso contrario → eteroschedasticità dei residui

In tale caso:

1) la significatività dei parametri non può essere valutata con il test *t*;
2) gli intervalli di confidenza delle previsioni non sono costanti nel tempo

Si utilizza il test di McLeod e Li: identico al test di Ljung-Box, ma fatto sui quadrati dei residui

Esempio

Calcolo di un AR(1) sui rendimenti del MIDEX

	coef	s.e.	t.stat	p.Value
ar1	0.1275	0.0232	5.5057	4.2E-08
c	0.0002	0.0003	0.7102	0.4777

NB.: il correlogramma dei quadrati delle differenze prime di un RW è sempre all'interno delle bande di oscillazione e il test di McLeod accetta l'omoschedisticità.

Test di McLeod e Li (test di LB sui residui al quadrato)

Lag	ACF	Q-Stat	Prob
0	1	47.45792268	5.62E-12
1	0.160512684	87.95175522	0
2	0.148228345	245.5297073	0
3	0.292325234	273.3035445	0
4	0.122692557	329.6219164	0
5	0.174665428	388.5213813	0
6	0.178574379	433.560216	0
7	0.156112813	502.2109109	0
8	0.192685432	569.4490541	0
9	0.190640708	594.2207078	0
10	0.115682214	605.998397	0
11	0.079744434	610.6993983	0
12	0.050367072	619.4490105	0
13	0.068695301	628.8538398	0
14	0.071201499	633.6522548	0
15	0.050844451	641.75689	0
16	0.066060569	648.0094295	0
17	0.058007567	656.5770122	0
18	0.067883835	660.8395965	0
19	0.047869008	660.9645351	0

I residui non sono omoschedastici

Controllo sulla normalità dei residui

-Indice di asimmetria di Fisher $$\gamma_1 = \frac{{}_M\mu_3}{\sigma^3}$$

con

$$ {}_M\mu_s = \frac{\sum_{t=1}^{T} (\hat{\varepsilon}_t)^s}{T} \quad \text{e} \quad \sigma = \sqrt{\frac{\sum_{t=1}^{T} (\hat{\varepsilon}_t)^2}{T}} $$

Asimmetria positiva o negativa se maggiore o minore di zero.

-Indice di curtosi $$\beta_2 = \frac{{}_M\mu_4}{{}_M\mu_2^2}$$

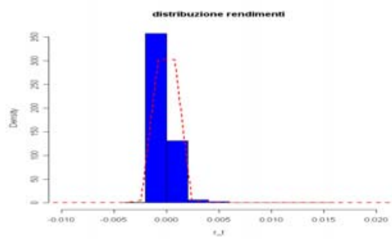
Normale → curtosi = 3
Platicurtica → curtosi < 3
Leptocurtica → curtosi > 3

- Test di Jarque-Bera $$JB = \frac{T}{6}\left[\gamma_1^2 + \frac{1}{4}(\beta_2 - 3)^2\right]$$

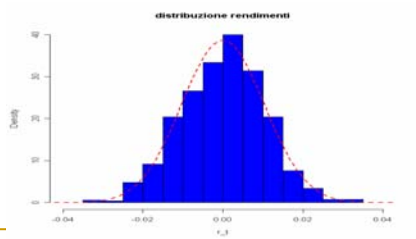
Sotto l'hp di normalità JB si distribuisce come una chi-quadrato con 2 gdl. Ipotesi nulla: normalità della serie.

Esempi

Istogramma dei residui di un ARMA(1,1) sul quadrato dei rendimenti del NASDAQ

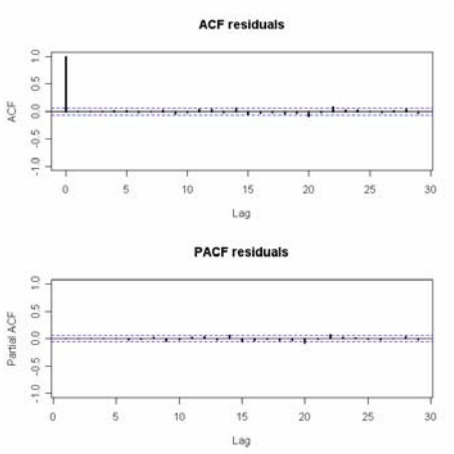


Istogramma dei residui di un modello ARMA(1,1) su una serie generata da un processo ARMA(1,1)

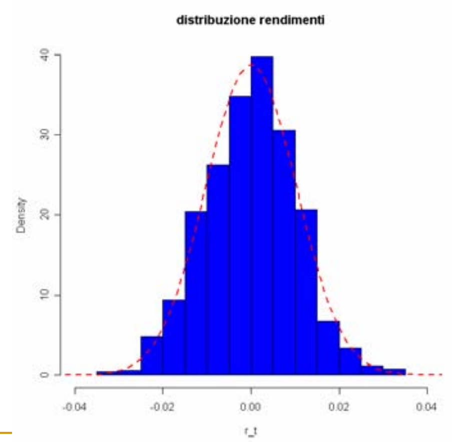


Adattamento del modello ai dati osservati

Correlogramma dei residui di un MA(5) su una serie generata da un processo ARMA(1,1)



Istogramma dei residui dello stesso modello



La scelta fra modelli alternativi

Come scegliere il modello migliore fra MA(5) e ARMA(1,1)?

-Criteri di scelta per *p* e *q*, che tengono conto del *trade-off* fra parsimonia (numero di parametri) e capacità previsiva del modello

Asymptotic Information Criterion (Akaike, 1974) $AIC(r) = T \ln \hat{\sigma}^2 + 2r$

con $r = p + q + 1$ e $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^2$

Bayesian Information Criterion (Schwarz, 1978) $BIC(r) = T \ln \hat{\sigma}^2 + r \ln T$

-Ciascun criterio assegna un costo all'introduzione di ogni nuovo parametro addizionale. I valori di *p* e di *q* che minimizzano i criteri rappresentano l'ordine del modello
-Il criterio BIC impone una penalità più alta di AIC per l'inclusione di nuove variabili

Esempio su ARMA(1,1) simulato (continua)

STIMA DI UN ARMA(1,1)

	coef	s.e.	t.stat	p.value
ar1	0.604	0.074	8.199	0.000
ma1	-0.789	0.057	-13.826	0.000
intercept	0.015	0.017	0.860	0.390

AIC 65.52 BIC 85.15

	coef	s.e.	t.stat	p.value
ma1	-0.175	0.032	-5.551	0.000
ma2	-0.138	0.032	-4.316	0.000
ma3	-0.062	0.033	-1.882	0.060
ma4	-0.019	0.031	-0.620	0.535
ma5	-0.044	0.032	-1.370	0.171
intercept	0.015	0.018	0.827	0.408

AIC 70.57 BIC 104.92

ATTENZIONE: R calcola AIC in modo diverso: -2*loglik+2*(r+1)

I valori minori di AIC e di BIC si hanno per il modello ARMA(1,1).