SERIE STORICHE E PROCESSI STOCASTICI

<u>Def</u> Un processo stocastico $\{X(t),\ t \in \mathcal{T}\}$ e' una collezione di v.c. indicizzate da un parametro t

Un p. s. e' detto a PARAMETRO DISCRETO se

$$t = 1, 2, \dots\dots$$

$\longrightarrow$ $X_1, X_2, \dots\dots$ oppure $\{X_t,\ t = 1, 2, \dots\}$

<u>Def</u> Un processo stocastico e' completamente noto se

- sono note le distribuzioni di ogni elemento

$$X_1, X_2, \dots$$

- sono note le distr. di ogni coppia di elementi

$$(X_1, X_2) \quad (X_1, X_3)$$
$$(X_2, X_5) \quad \dots$$

$\dots$

- sono note le distr. di ogni terna

$\dots$

NB

Fissiamo un istante temporale

$t=5 \longrightarrow X_5$ e' una v.c. (assume diversi valori con diverse prob.)

$\downarrow$

x_5 e' uno dei valori assunti da X_5

Così

$X_1, X_2, \ldots, X_n, \ldots \longrightarrow x_1, x_2, \ldots, x_n, \ldots$

e' uno dei possibili insiemi di realizzazioni del p.s.

$\longrightarrow$ REALIZZAZIONE o TRAIETTORIA del processo

Una serie storica può essere pensata come la parte finita di una realizzazione di un processo stocastico.

Dalla serie storica si cerca di apprendere qualcosa sull'intero processo che l'ha generata

PARTICOLARI PROCESSI STOCASTICI

① P. "WHITE NOISE"

$A_1, A_2, \dots$ t.c. (NB) sul libro $A \to a$

1) $E(A_t) = 0$

2) $E(A_t^2) = \sigma_A^2$

3) $\mathrm{cov}(A_t, A_s) = 0 \quad \forall t \neq s$

Osservazioni: - medie e var. costanti
- componenti incorrelate (di solito è un'ipotesi poco realistica)

② P. GAUSSIANI

$G_1, G_2, \dots$ t.c.

$G_i \sim \text{Normale}(\mu_i, \sigma_i^2) \quad \mathrm{cov}(G_i, G_j) = \sigma_{ij}$

NB: - Conoscere i parametri significa conoscere l'intero processo
- se $\sigma_{ij} = 0$ il processo è a componenti indipendenti
- caso particolare: WN gaussiano

③ P. STAZIONARI IN SENSO FORTE

un p.s. e' stazionario in senso forte se la distribuzione di

$$X_{t_1}, X_{t_2}, X_{t_3}, \dots, X_{t_k}$$

e' sempre la stessa qualunque k-upla di valori $t_1, \dots, t_k$ si scelga

<u>Es</u>

$$(X_2, X_7, X_9) \overset{\mathcal{L}}{=} (X_3, X_{20}, X_{21})$$

$$X_1 \overset{\mathcal{L}}{=} X_7$$

$$\vdots$$

e' una condizione molto restrittiva

④ P.S. STAZIONARI IN SENSO DEBOLE

1) $E(X_t) = \mu \qquad \forall t$

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

3) $cov(X_t, X_s) = \gamma(|s-t|)$

NB:

- staz. senso forte $\Rightarrow$ staz. senso debole

- staz. senso debole + gaussiano $\Rightarrow$ staz. senso forte

- $Cov(X_t, X_{t+k}) = \gamma(|t+k-t|) = \gamma(k)$

  $Cov(X_{t-k}, X_t) = \gamma(|t-t+k|) = \gamma(k)$

  la covarianza e' funzione solo del "lag"

- si possono considerare, senza perdite di generalita', p.s. stazionari a media nulla:

$$X_t \text{ staz.} \longrightarrow Z_t = X_t - \mu$$

$$E(Z_t) = 0$$
$$V(Z_t) = \sigma^2$$
$$cov(Z_t, Z_s) = \gamma(|t-s|)$$

⑤ P.S. INVERTIBILI

Un p.s. e' INVERTIBILE se esiste una funzione $h(\cdot)$ t.c.

$$X_t = h(X_{t-1}, X_{t-2}, \dots) + A_t$$

essendo $A_t \sim WN$

(si può derivare in funzione del "passato")

FUNZIONE DI AUTOCORRELAZIONE

Una serie storica e' pensata come la realizzazione di un processo stocastico. In ogni serie storica c'e' un legame tra il valore corrente ed il "passato"

$$X_t \longleftarrow X_{t-1} \longleftarrow X_{t-2} \longleftarrow \cdots$$

$$X_t \longleftarrow X_{t-12} \longleftarrow X_{t-24} \longleftarrow \cdots$$

Per studiare questi legami si puo' considerare la fn di autocorrelazione

X_1	-
X_2	X_1
X_3	X_2
X_4	X_3
X_5	X_4
$\vdots$	$\vdots$
X_t	X_{t-1}
$\vdots$	$\vdots$

$$Cov(X_t, X_{t-1})$$

$$\frac{Cov(X_t, X_{t-1})}{\sqrt{Var(X_t)\,Var(X_{t-1})}}$$

quanto legame lineare lineare c'e' con il valore precedente

$$\begin{array}{cc}
X_1 & - \\
X_2 & - \\
X_3 & X_1 \\
X_4 & X_2 \\
\vdots & \vdots \\
X_t & X_{t-2} \\
\vdots & \vdots
\end{array}$$

$Cov(X_t, X_{t-2})$

$$\frac{Cov(X_t, X_{t-2})}{\sqrt{Var(X_t)\, Var(X_{t-2})}}$$

legame con due valori passati

$Cov(X_t, X_{t-12})$

legame con gli anni passati (dati annuali)

$\vdots$

Se il processo e' stazionario

$$Cov(X_t, X_{t-1}) = \gamma(|t-1-t|) = \gamma(1)$$

$$Var(X_t) = \gamma(0)$$
$$Var(X_{t-1}) = \gamma(0)$$

$$\frac{Cov(X_t, X_{t-1})}{\sqrt{Var(X_t)\, Var(X_{t-1})}} = \frac{\gamma(1)}{\gamma(0)} = \rho(1)$$

$$\frac{cov(X_t, X_{t-2})}{\sqrt{var(X_t)\, var(X_{t-2})}} = \frac{\gamma(2)}{\gamma(0)} = \rho(2)$$

$$\frac{cov(X_t, X_{t-12})}{\sqrt{var(X_t)\, var(X_{t-12})}} = \frac{\gamma(12)}{\gamma(0)} = \rho(12)$$

In generale

$$\frac{cov(X_t, X_{t-k})}{\sqrt{var(X_t)\, var(X_{t-k})}} = \frac{\gamma(k)}{\gamma(0)} = \underbrace{\rho(k)}_{\text{FUNZIONE DI AUTO CORRELAZ. DEL PROCESSO}}$$

Proprietà della fn. di autocorrelazione

1) $\rho(0) = 1$ infatti $\rho(0) = \dfrac{\gamma(0)}{\gamma(0)} = 1$

2) $\rho(-k) = \rho(k)$

infatti

$$\gamma(-k) = \operatorname{Cov}(X_t, X_{t+k}) =$$
$$= \operatorname{Cov}(X_{t+k}, X_t) = \gamma(k)$$

3) $|\rho(k)| \leq 1$

4) $X_t \longrightarrow a + b X_t = Y_t$

$$|\rho_Y(k)| = |\rho_X(k)|$$

5) la matrice

$$P_{(m)} = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(m-1) \\ \rho(1) & 1 & \rho(2) & \dots & \rho(m-2) \\ \rho(2) & \rho(1) & 1 & \dots & \rho(m-3) \\ \vdots & & & & \vdots \\ \rho(m-1) & \rho(m-2) & \rho(m-3) & \dots & 1 \end{bmatrix}$$

e' semidefinita positiva
(minori principali ≥ 0)

Problema

La fu di autocorrelazione e' nota solo se ~~~~ e' noto il p.s. - Noi conosciamo solo una realizzazione del p.s. - Come possiamo stimare l'intera funzione?

In realta' il problema si pone già per stimare la media μ del processo

$$E(X_t) = \mu$$

noi disponiamo di

$$\begin{array}{cccc} X_1, & X_2, & \dots, & X_N, \dots \\ \downarrow & \downarrow & \dots & \downarrow \\ x_1 & x_2 & & x_N \end{array}$$

ogni elemento della serie storica e' la realizzazione di un campione unitario.
Bisognerebbe disporre di un campione di una certa ampiezza per ciascuna v.c. del processo

Se il processo e' regolare, tuttavia, e' possibile pensare ad una stima "temporale"

$$\bar{x}_N = \frac{1}{N} \sum_{t=1}^{N} x_t$$

se questa stima ~~~~ e' consistente il processo si dice ERGODICO
(meglio: se lo stimatore corrispondente e' -10-

Consistente in media quadratica

$$\lim_{N \to \infty} E(\bar{X}_N - \mu)^2 = 0$$

$$\left(NB \Rightarrow \lim_{T \to \infty} Pr\{ |\bar{X}_T - \mu| < \varepsilon \} = 1 \right)$$

Considereremo solo processi stazionari ed ergodici

Stima di $\rho(k) \longrightarrow \hat{\rho}(k) = \dfrac{\hat{\gamma}(k)}{\hat{\gamma}(0)}$

con

$$\hat{\gamma}(k) = \frac{1}{N} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})$$

$$\hat{\gamma}(0) = \frac{1}{N} \sum_{t=1}^{N} (x_t - \bar{x})^2$$

dunque

$$\hat{\rho}(k) = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{N} (x_t - \bar{x})^2}$$

Naturalmente la stima può essere affetta da errore campionario. Per sapere qualcosa, bisogna conoscere le proprietà dello stimatore corrispondente.

Lo stimatore prescelto per $\rho(k)$

- è distorto
- è tuttavia più efficiente rispetto ad altri stimatori non distorti

Lo stimatore prescelto per $\rho(k)$

- è asintoticamente normale
- ha varianza approssimativamente pari ad $\frac{1}{N}$

$\Rightarrow$ I.C. per $\rho(k)$ al 95%

$$\left(\hat{\rho}(k) - 1{,}96 \cdot \frac{1}{\sqrt{N}} \; ; \; \hat{\rho}(k) + 1{,}96 \, \frac{1}{\sqrt{N}} \right)$$

oppure

$$\left(\hat{\rho}(k) - \frac{2}{\sqrt{N}} \; ; \; \hat{\rho}(k) + \frac{2}{\sqrt{N}} \right)$$

"banda" di confidenza

oppure, volendo testare $H_0: \rho(k) = 0$

Rifiuto H_0 se

$$\hat{\rho}(k) \leq -\frac{2}{\sqrt{N}}$$

$$\hat{\rho}(k) \geq \frac{2}{\sqrt{N}}$$

Esempio

1. storica 1, 2, 7, 3, 5, 2, 0, 1, 4, 5

x_t	x_{t+1}	$(x_t - \bar{x})$	$(x_t - \bar{x})^2$	$(x_{t+1} - \bar{x})$	$(x_{t+1} - \bar{x})(x_t - \bar{x})$
1	2	-2	4	-1	2
2	7	-1	1	4	-4
7	3	4	16	0	0
3	5	0	0	2	0
5	2	2	4	-1	-2
2	0	-1	1	-3	3
0	1	-3	9	-2	6
1	4	-2	4	1	-2
4	5	1	1	2	2
5	—	2	4	—	—
30			44		5

$$\bar{x} = \frac{30}{10} = 3$$

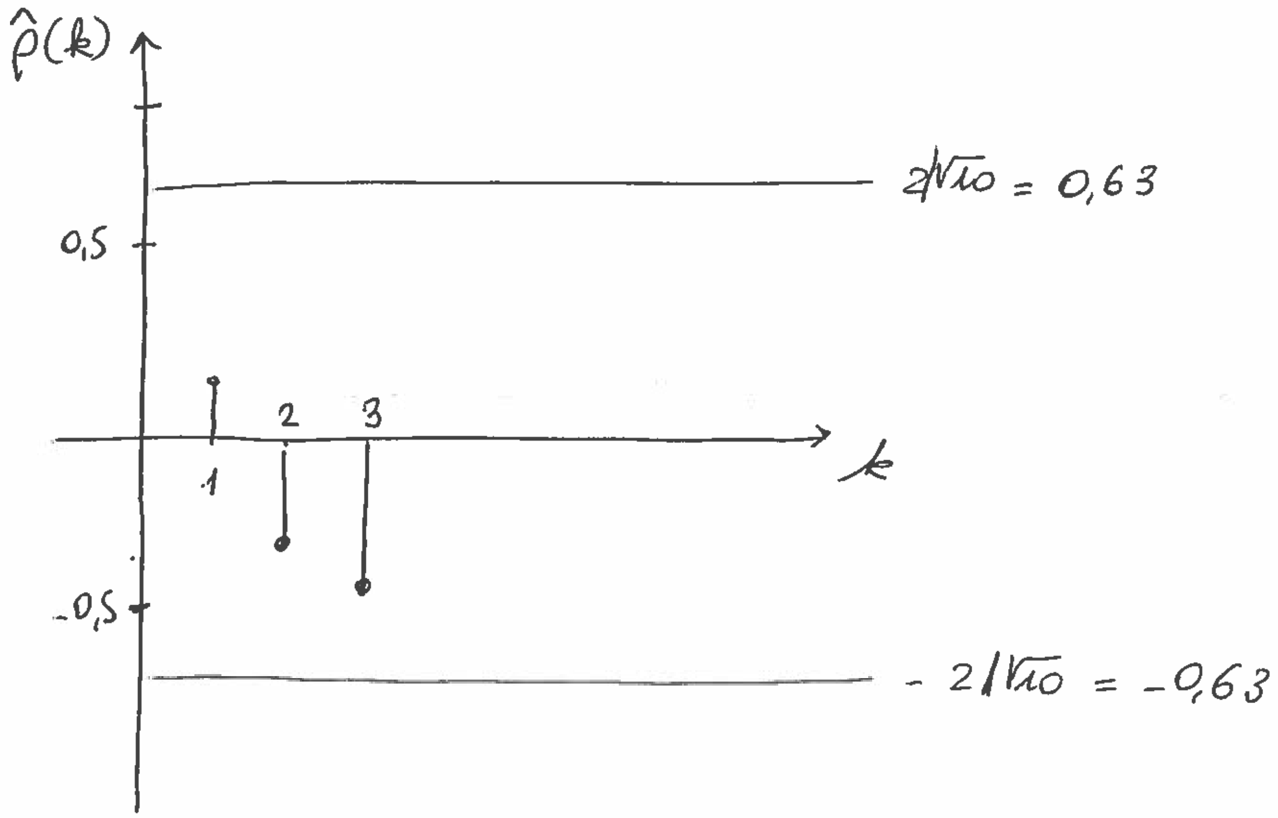
$$\hat{\gamma}(0) = \frac{44}{10} = 4,4 \qquad \hat{\gamma}(1) = \frac{5}{10} = 0,5$$

$$\hat{\rho}(1) = \frac{0,5}{4,4} = 0,1136$$

e così si ottiene

$$\hat{\rho}(2) = -0{,}25$$

$$\hat{\rho}(3) = -0{,}38$$



<u>Correlogramma</u>

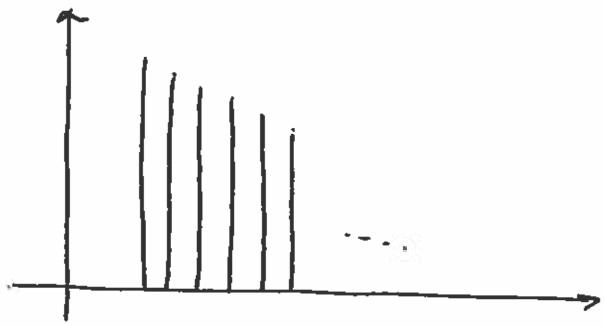
- solo valori positivi di k
- valori di k da 1 fino a $\frac{N}{4}$ (in generale)

Uso delle f.ni di autocorrelazione (e del correlogramma

① serie con trend

$$X_t \leftarrow X_{t-1} \leftarrow X_{t-2}$$

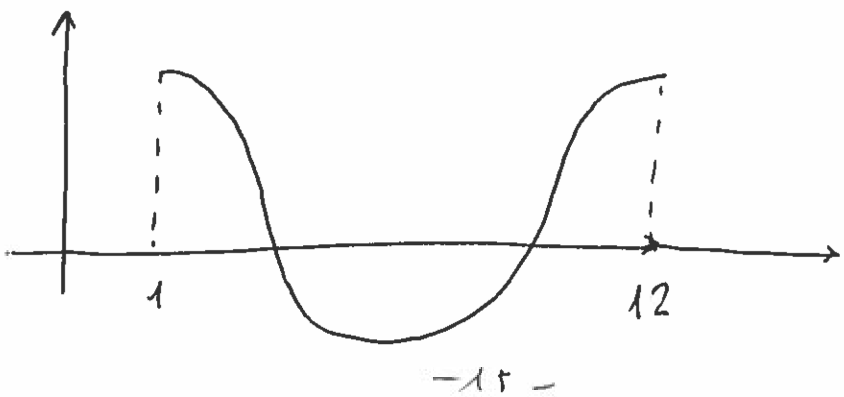
elevata correlazione al lag 1
correlazione positiva al lag 2
correlazione si smorza

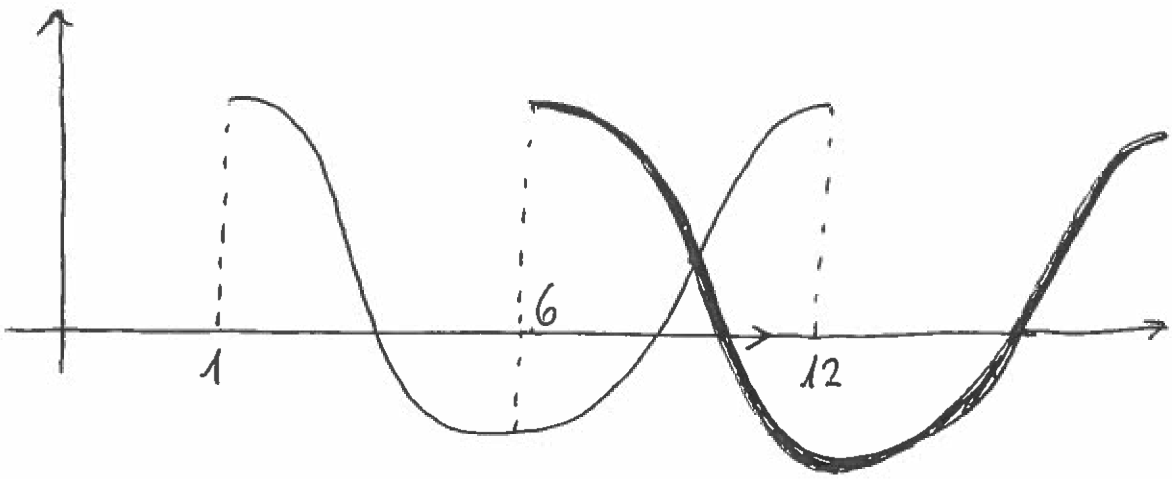


② serie con stagionalità

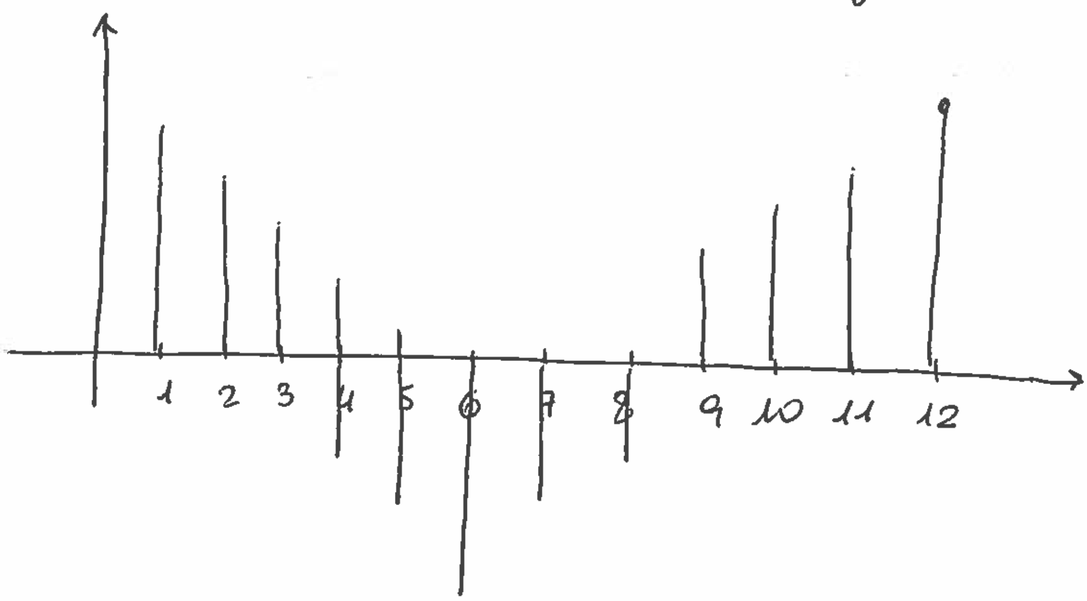
$$X_t \leftarrow X_{t-12}$$

$$X_t \leftarrow X_{t-6} \quad ??$$

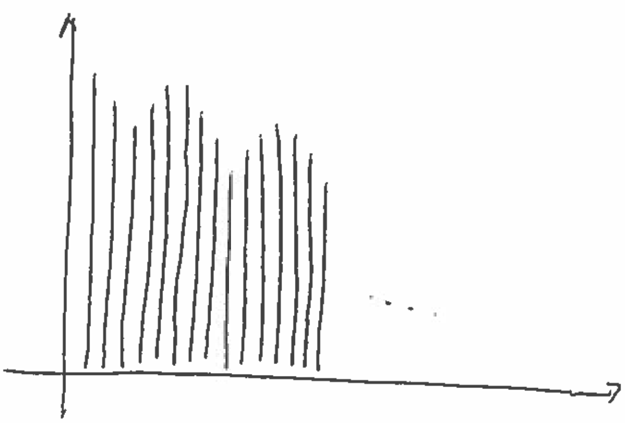




X_t e X_{t-6} $\longrightarrow$ correlazione negativa



③ trend + stagionalita'



FUNZ. DI AUTOCORRELAZIONE PARZIALE

$$\begin{tikzcd} X_t & X_{t-1} \arrow[l] & X_{t-2} \arrow[l] & X_{t-3} \arrow[l] \arrow[lll, bend left, dashed] \end{tikzcd}$$

$\rho(3)$ studia il legame tra X_t e X_{t-3}
che e' influenzato dal fatto
che, se X_{t-3} varia, anche
X_{t-1} e X_{t-2} variano ed hanno
influenza su X_t

$\pi(3)$ legame tra X_t e X_{t-3}
mantenendo costanti X_{t-2} e X_{t-1}
(coeff. di correlazione parziale)

$\pi(k)$ FUNZ. DI AUTOCORRELAZIONE PARZIALE:
corr. tra X_t e X_{t-k} mantenendo
costanti le v.c. intermedie
X_{t-1} ..... X_{t-k+1}

Si dimostra che

$$\pi(k) = \frac{\det(P^*_{(k)})}{\det(P_{(k)})}$$

con

$$P_{(k)} = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \cdots & \rho(k-1) \\ \rho(1) & 1 & \rho(1) & \cdots & \rho(k-2) \\ \rho(2) & \rho(1) & 1 & \cdots & \rho(k-3) \\ \vdots & & & & \vdots \\ \rho(k-1) & \rho(k-2) & \rho(k-3) & \cdots & 1 \end{bmatrix}$$

$$P^*_{(k)} = \begin{bmatrix} & & \rho(1) \\ & & \rho(2) \\ & \text{idem} & \vdots \\ & & \\ & & \rho(k) \end{bmatrix}$$

Pertanto, con alcuni passaggi, si ottiene ad es.

$$\pi(1) = \rho(1)$$

$$\pi(2) = \frac{\rho(2) - \rho^2(1)}{1 - \rho^2(1)}$$

$$\pi(3) = \frac{\rho(3)\left[1 - \rho^2(1)\right] + \rho(1)\left[\rho^2(1) + \rho^2(2) - 2\rho(2)\right]}{\left[1 - \rho(2)\right]\left[1 + \rho(2) - 2\rho^2(1)\right]}$$

$$\vdots$$

Stima della funz. di autocorrelazione parziale

Basta sostituire $\rho(k)$ con $\hat{\rho}(k)$:

$$\hat{\pi}(k) = \frac{\det\left(\hat{P}^*_{(k)}\right)}{\det\left(\hat{P}_{(k)}\right)}$$

Varianza stimatore e bande di confidenza $\longrightarrow$ come per $\rho(k)$

Uso

Serve ad identificare particolari tipi di modelli per il p. s. (ARIMA)

MODELLI ARIMA

Cerchiamo un MODELLO per il processo stocastico. Così come avviene per le popolazioni, ipotizzare una certa distribuzione, permette di concentrarci solo sui parametri che la caratterizzano.

Nel cercare un modello, ci limiteremo a considerare p.s. stazionari ed invertibili (due proprietà altamente desiderabili per la nostra analisi)

<u>Teorema di Wold</u>

Ogni processo stocastico stazionario X_t si può scomporre come

$$X_t = Z_t + V_t$$

con

$$Z_t = A_t + \psi_1 A_{t-1} + \psi_2 A_{t-2} + \dots \quad \left(\sum \psi_j^2 < \infty \right)$$

$$V_t = \mu + \sum_{j=1}^{\infty} \{ \alpha_j \cos \lambda_j t + \beta_j \sin \lambda_j t \}$$

e $Cov(Z_t, V_t) = 0$

(NB) • A_t e' il processo White noise

• $A_t \sim WN(0; \sigma_a^2)$

• $\alpha_1, \dots \quad \beta_1, \beta_2, \dots$ sono v.c. con

$$E(\alpha_j) = E(\beta_j) = 0$$

$$cov(\alpha_j, \beta_j) = 0$$

• $\lambda_1, \lambda_2, \dots$ sono numeri reali t.c.
$0 \leq \lambda_j \leq \pi$ (frequenza angolare deterministica)

Z_t e' detta componente <u>non deterministica</u> o <u>puramente stocastica</u>

→ ha una struttura decisamente casuale, pur essendo combinazione casuale di strutture semplici (w.n.)

V_t e' detta componente <u>deterministica</u>

→ non perche' non contenga elementi casuali, ma perche' ha una struttura "stabile"; e' una c.l. di funz. periodiche le cui caratteristiche possono essere stimate:

(ampiezza e fase)

$$\min_{\alpha \beta} \sum (x_t - v_t)^2$$

metodo dei minimi quadrati per stimare α_j e β_j

La parte più interessante (più difficile da analizzare) è la componente Z_t:

$$Z_t = A_t + \psi_1 A_{t-1} + \psi_2 A_{t-2} + \dots$$

Si tratta di un processo stazionario, poiché si può dimostrare che

- $E(Z_t) = 0$
- $Var(Z_t) = \sigma_a^2 \sum_{j=1}^{\infty} \psi_j^2 < \infty$
- $Cov(Z_t, Z_{t-k}) = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k}$

Poiché $\sum \psi_j^2 < \infty$, gli ψ_j saranno molto piccoli da un certo punto in poi. Pertanto potremmo porre:

$$\psi_j = -\theta_j \qquad j = 1, \dots, q$$

$$\psi_j = 0 \qquad j = q+1, q+2, \dots$$

$$\Rightarrow Z_t = A_t - \theta_1 A_{t-1} - \theta_2 A_{t-2} - \dots - \theta_q A_{t-q}$$

PROCESSO "MEDIA MOBILE" DI ORD. q
(MA(q))

Il processo MA(q) si scrive spesso usando l' "operatore backward":

$$B X_t = X_{t-1}$$

$$X_{t-2} = B(B X_t) = B^2 X_t$$

$$X_{t-k} = B^k X_t$$

Si ha

$$Z_t = A_t - \theta_1 B A_t - \theta_2 B^2 A_t - \dots - \theta_q B^q A_t =$$

$$= (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) A_t =$$

$$= \theta(B) A_t$$

Il processo $Z_t \sim MA(q)$ e' sicuramente stazionario. Si puo' dim. che e' anche invertibile se

le radici di $\theta(B) = 0$
sono tutte in modulo > 1

NB: $\theta(B) = 0$

$$1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q = 0$$

Esempio

$Z_t \sim MA(1)$

$$Z_t = A_t - \theta A_{t-1}$$

$$\theta(B) = 0 \implies 1 - \theta B = 0$$

$$\implies B = \frac{1}{\theta}$$

$$\implies \left|\frac{1}{\theta}\right| > 1 \implies |\theta| < 1$$

$$-1 < \theta < 1$$

Se MA(q) e' invertibile, si puo' scrivere in funzione del "passato" – Proviamo a farlo per un MA(1)

$$Z_t = A_t - \theta A_{t-1} = (1 - \theta B) A_t$$

$$A_t = \frac{1}{1 - \theta B} Z_t$$

$$A_t = \sum_{i=0}^{\infty} (\theta B)^i Z_t$$

$$A_t = (1 + \theta B + \theta^2 B^2 + \theta^3 B^3 + \dots) Z_t$$

$$A_t = Z_t + \theta Z_{t-1} + \theta^2 Z_{t-2} + \dots$$

$$\implies Z_t = \cancel{Z} - \theta Z_{t-1} - \theta^2 Z_{t-2} - \dots + A_t$$

Cambiando simboli

$$Z_t = \pi_1 Z_{t-1} + \pi_2 Z_{t-2} + \dots + A_t$$

questa struttura e' detta AUTOREGRESSIVA: e' come un modello lineare in cui Z_t viene regredito su se stesso ad istanti temporali diversi

Anche in questo caso, si puo' dimostrare che i pesi π_j sono molto piccoli da un certo punto in poi_ Pertanto si puo' approssimare:

$$\pi_j = \phi_j \qquad j = 1, \dots, p$$
$$\pi_j = 0 \qquad j = p+1, \dots$$

$$\longrightarrow \quad Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + A_t$$

questo e' il PROCESSO AUTOREGRESSIVO di ordine p ($AR(p)$)

Anche AR(p) si può scrivere usando l'operatore backward

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + A_t$$

$$= \phi_1 B Z_t + \phi_2 B^2 Z_t + \dots + \phi_p B^p Z_t + A_t$$

$$Z_t - \phi_1 B Z_t - \phi_2 B^2 Z_t - \dots - \phi_p B^p Z_t = A_t$$

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Z_t = A_t$$

$$\phi(B) Z_t = A_t$$

Il processo AR(p) è sicuramente invertibile, ma non necessariamente stazionario. Si può dim. che è stazionario se

le radici dell'equazione
$\phi(B)=0$ sono in modulo
maggiori di 1

un processo AR(p) è molto "simile" ad un processo media mobile. Infatti un AR(p) si può scrivere come un MA(∞):

ad es. nel caso AR(1)

$$(1-\phi B) Z_t = A_t$$

$$Z_t = A_t \frac{1}{1-\phi B}$$

$$Z_t = \sum_{j=0}^{\infty} (\phi B)^j A_t$$

$$= (1 + \phi B + \phi^2 B^2 + \dots) A_t$$

$$= A_t + \phi A_{t-1} + \phi^2 A_{t-2} + \dots$$

processo MA(∞)

Dunque abbiamo a disposizione due modelli:

MA(q) $\qquad Z_t = \theta(B) A_t$

AR(p) $\qquad \phi(B) Z_t = A_t$

sono due facce della stessa medaglia, ma considerano due logiche leggermente diverse

AR $\longrightarrow$ spiegano il presente in funzione del passato

MA $\longrightarrow$ spiegano il presente in funzione di una serie di "impulsi" casuali

Si possono pertanto "fondere" le due logiche:

$$\phi(B) Z_t = \theta(B) A_t$$

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Z_t = (1 - \theta_1 B - \theta_2 B - \dots - \theta^q B^q) A_t$$

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + A_t - \theta_1 A_{t-1} - \dots - \theta_q A_{t-q}$$

MODELLO ARMA(p, q)

MODELLI ARIMA

L'ipotesi di partenza che abbiamo fatto, e' che il processo Z_t su cui lavoriamo sia stazionario. Esistono, tuttavia, molte possibilita' di allontanamento da questa ipotesi:

non stazionarieta'
- in media
- in varianza
- in autocovarianza

(vedi grafici)

Per correggere la non-stazionarieta' in media e' possibile estendere la classe dei modelli ARMA:

Consideriamo un processo Z_t dotato di trend lineare, dunque non stazionario in media

$$Z_t = \beta_0 + \beta_1 t + \underbrace{W_t}_{\text{parte stazionaria del modello}}$$

Consideriamo ora l'incremento del processo:

$$Z_t - Z_{t-1} = \beta_0 + \beta_1 t + W_t - (\beta_0 + \beta_1 (t-1) + W_{t-1}) =$$

$$= \beta_1 + \underbrace{W_t - W_{t-1}}$$

e' un[illegible] stazionario

L'operazione che porta ad ottenere

$$Z_t \longrightarrow D_t = Z_t - Z_{t-1}$$

è detta di DIFFERENZIAZIONE di ord 1

$$D_t = Z_t - Z_{t-1} = (1-B)Z_t = \nabla Z_t$$

differenza prima di ordine 1

Se il trend fosse quadratico?

$$Z_t = \beta_0 + \beta_1 t + \beta_2 t^2 + W_t$$

differenziamo

$$D_t = Z_t - Z_{t-1} = \beta_0 + \beta_1 t + \beta_2 t^2 + W_t - (\beta_0 + \beta_1 (t-1) + \beta_2 (t-1)^2 + W_{t-1})$$

$$= \beta_1 - \beta_2 + 2t\beta_2 + W_t - W_{t-1}$$

differenziamo ancora

$$E_t = D_t - D_{t-1} = \beta_1 - \beta_2 + 2t\beta_2 + W_t - W_{t-1} - (\beta_1 - \beta_2 - 2(t-1)\beta_2 + W_{t-1} - W_{t-2}) =$$

$$= 2\beta_2 + W_t - 2W_{t-1} + W_{t-2}$$ stazionario

qui abbiamo

$$Z_t \longrightarrow E_t = D_t - D_{t-1} =$$

$$= Z_t - Z_{t-1} - (Z_{t-1} - Z_{t-2}) =$$

$$= Z_t - 2Z_{t-1} + Z_{t-2}$$

$$E_t = (1-B) D_t = (1-B)(1-B) Z_t =$$

$$= (1-B)^2 Z_t = \nabla^2 Z_t$$

differenziazione di ordine 2

In generale

$$\nabla^d Z_t = (1-B)^d Z_t \quad \text{diff. di ord. } d$$

Al processo differenziato si possono applicare i modelli ARMA!

$$U_t = \nabla^d Z_t$$

$$\phi(B) U_t = \theta(B) A_t$$

$$\phi(B) \nabla^d Z_t = \theta(B) A_t$$

PROCESSO Autoregressivo Media Mobile INTEGRATO

(ARIMA(p, d, q))

$$(1-\phi_1 B-\phi_2 B^2-\dots-\phi_p B^p)(1-B)^d Z_t=(1-\theta_1 B-\theta_2 B_2-\dots-\theta_q B^q)A_t$$

<u>Ad es.</u> ARIMA (1,1,2)

$$(1-\phi B)(1-B)Z_t=(1-\theta_1 B-\theta_2 B^2)A_t$$

$$(1-\phi B)(Z_t-Z_{t-1})=A_t-\theta_1 A_{t-1}-\theta_2 A_{t-2}$$

$$Z_t-Z_{t-1}-\phi Z_{t-1}+\phi Z_{t-2}=A_t-\theta_1 A_{t-1}-\theta_2 A_{t-}$$

$$Z_t=(1+\phi)Z_{t-1}-\phi Z_{t-2}+A_t-\theta_1 A_{t-1}-\theta_2 A_{t-}$$

IDENTIFICAZIONE DI UN MODELLO ARIMA

Significa in sostanza determinare i valori di p, d, q -
Possiamo tuttavia delineare alcuni passaggi necessari a questo scopo -

1) Analisi preliminari
   - grafici
   - eventuali trasformazioni
   - presenza outlier

2) Identificazione del modello ARIMA
   - fu autocorrelazione
   - valori di p, d, q

3) Stima dei parametri

4) Verifica del modello
   - analisi dei residui (white noise?)

5) previsione (se il modello e' accettato)

(1) Analisi preliminari

E' necessario controllare che le ipotesi dei modelli ARIMA siano soddisfatte. In primo luogo la stazionarietà.

Non stazionarietà in media :
- stima di un trend polinomiale col metodo dei minimi quadrati
- differenziazione (più usata)
- var. esplicative ?

Non stazionarietà in varianza : - trasf. logaritmica

Outliers : - tecniche di perequazione (medie mobili, interpolaz. lineare)

② Identificazione

- Stima della fn di autocorrelazione globale/parziale

  $\longrightarrow$ bande di confidenza !!

- Analisi secondo il seguente schema:

Modello	Fn autoc. globale	Fn autoc. parziale
MA(q)	si annulla dopo il lag q	decade in modo esponenziale
AR(p)	decade in modo esponenziale	si annulla dopo il lag p
ARMA(p,q)	decade in modo esp. dopo il lag p-q	decade in modo esp. dopo il lag. q-p

(Vedi schema SPSS)

③ Stima

Metodo della massima verosimiglianza

→ ricorso a tecniche di ottimizzazione numerica

Metodo di MARQUARDT
(valori iniziali)

④ Verifica del modello

Bisogna controllare che i residui siano realmente dei White Noise

- analisi grafiche (outlier, pattern)
- analisi di "normalità"
- analisi di autocorrelazione dei residui

→ eventuale modifica del modello *

⑤ previsione

(solo con il calcolatore)

* Spesso la verifica avviene anche con la tecnica di <u>sovraparametrizzazione</u>

si stima un modello più complesso e si vede se tutti i suoi parametri sono significativi

② Identificazione

- Stima della fn di autocorrelazione globale/parziale

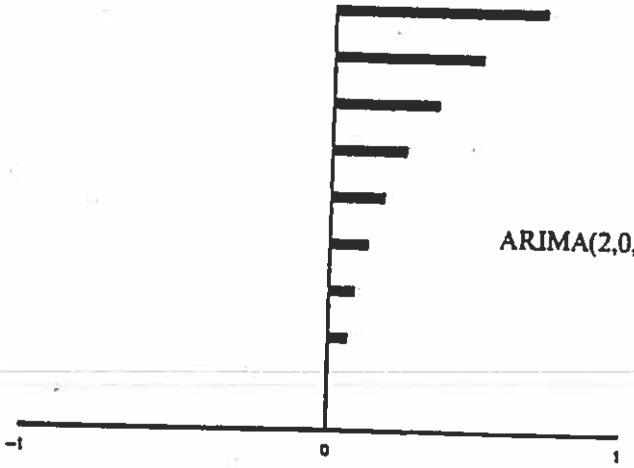
  → bande di confidenza !!

- Analisi secondo il seguente schema:

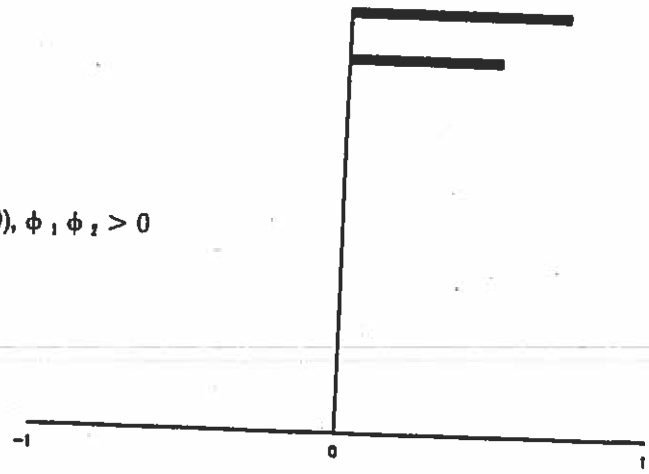
Modello	Fn autoc. globale	Fn autoc. parziale
MA(q)	si annulla dopo il lag q	decade in modo esponenziale
AR(p)	decade in modo esponenziale	si annulla dopo il lag p
ARMA(p,q)	decade in modo esp. dopo il lag p-q	decade in modo esp. dopo il lag q-p

(Vedi schema SPSS)

ACF

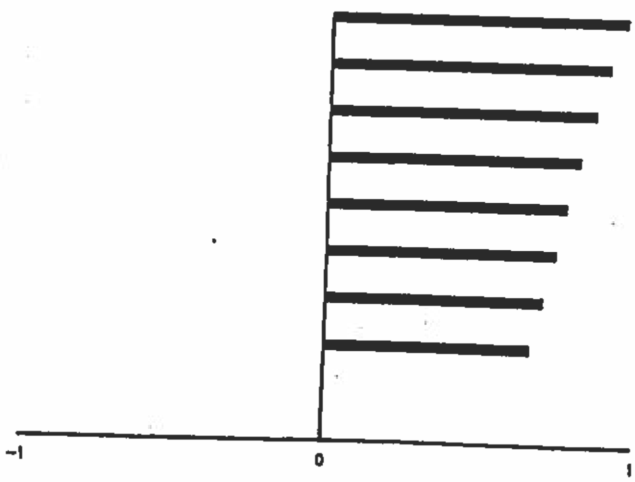


PACF

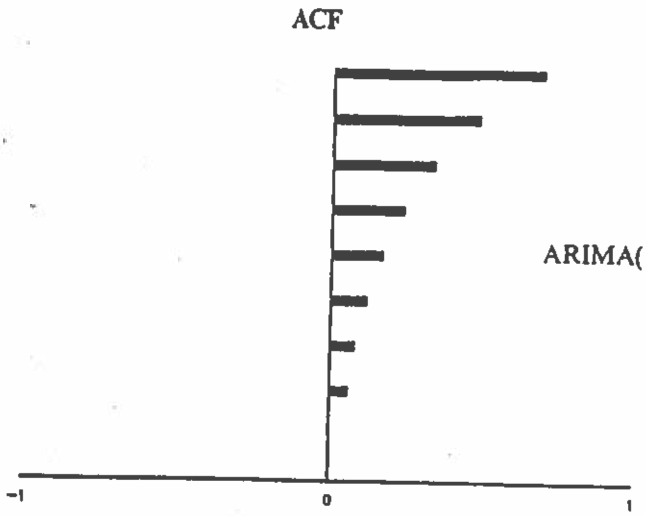


ARIMA(2,0,0), $\phi_1 \phi_2 > 0$

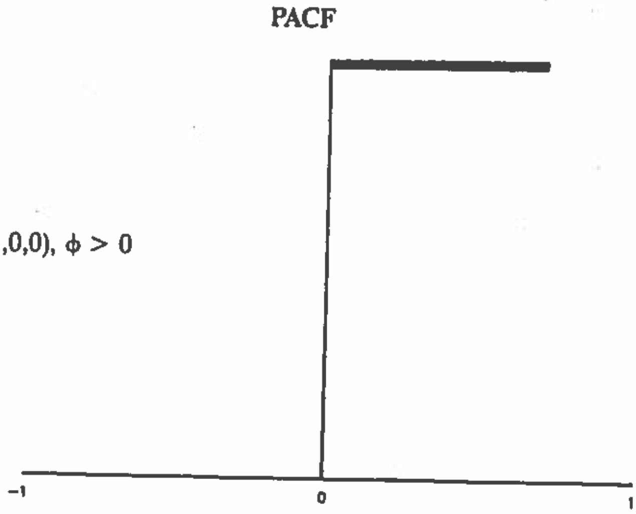
ACF

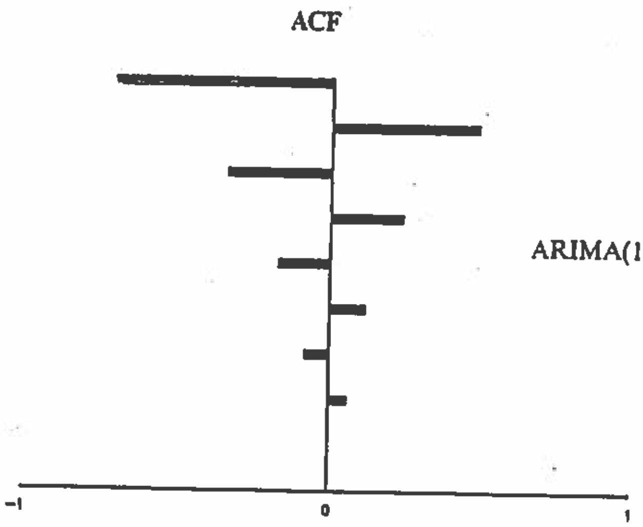


ARIMA(0, 1, 0) (integrated series)

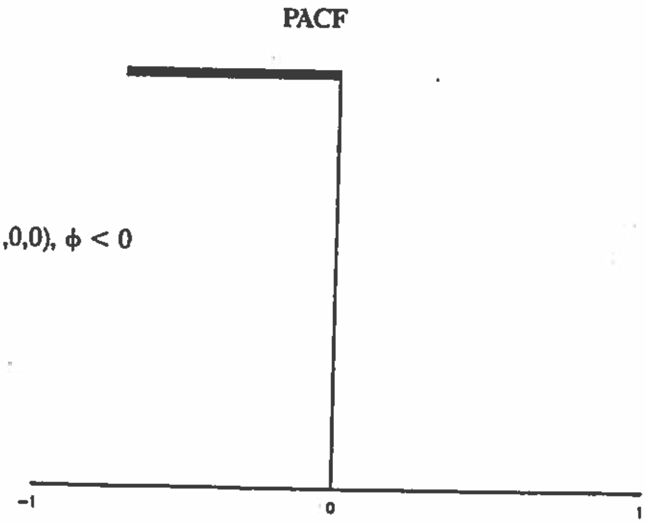


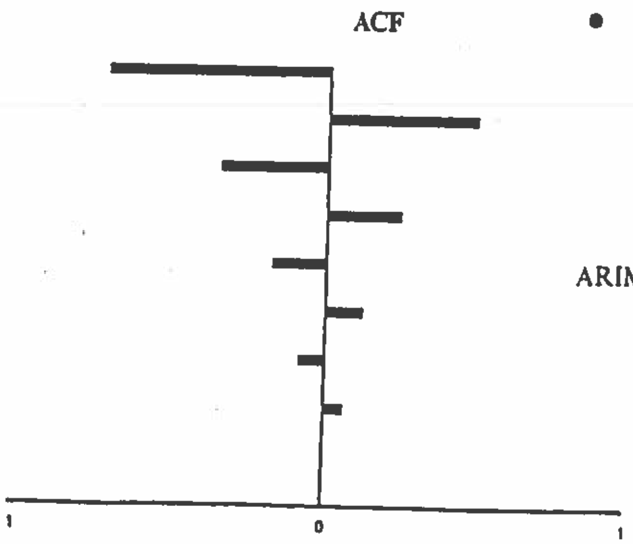
ARIMA(1,0,0), $\phi > 0$



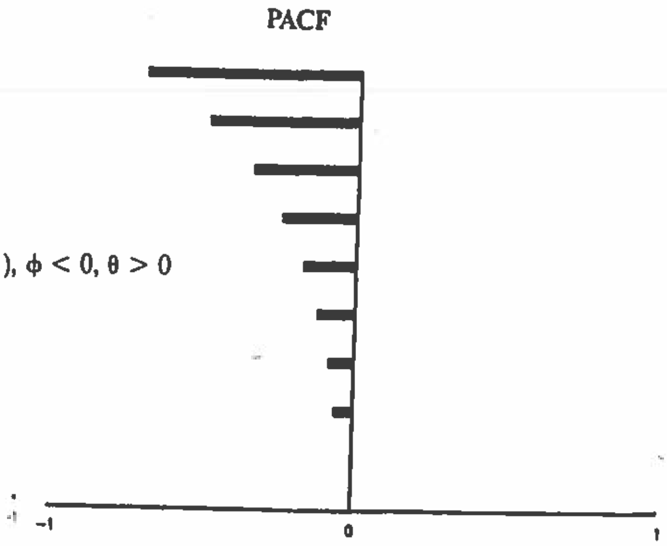


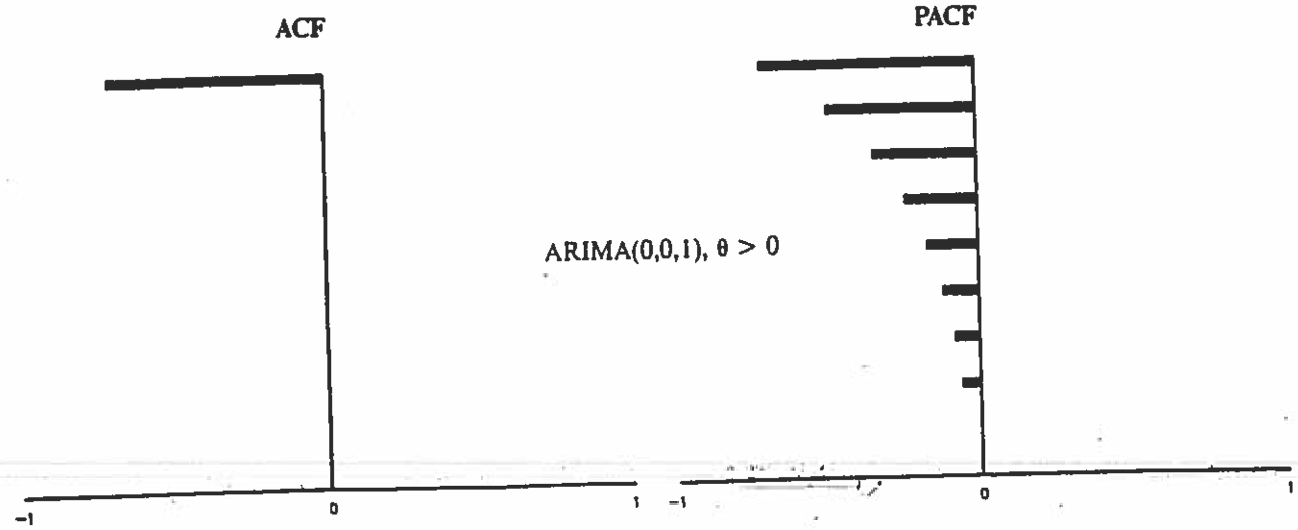
ARIMA(1,0,0), $\phi < 0$

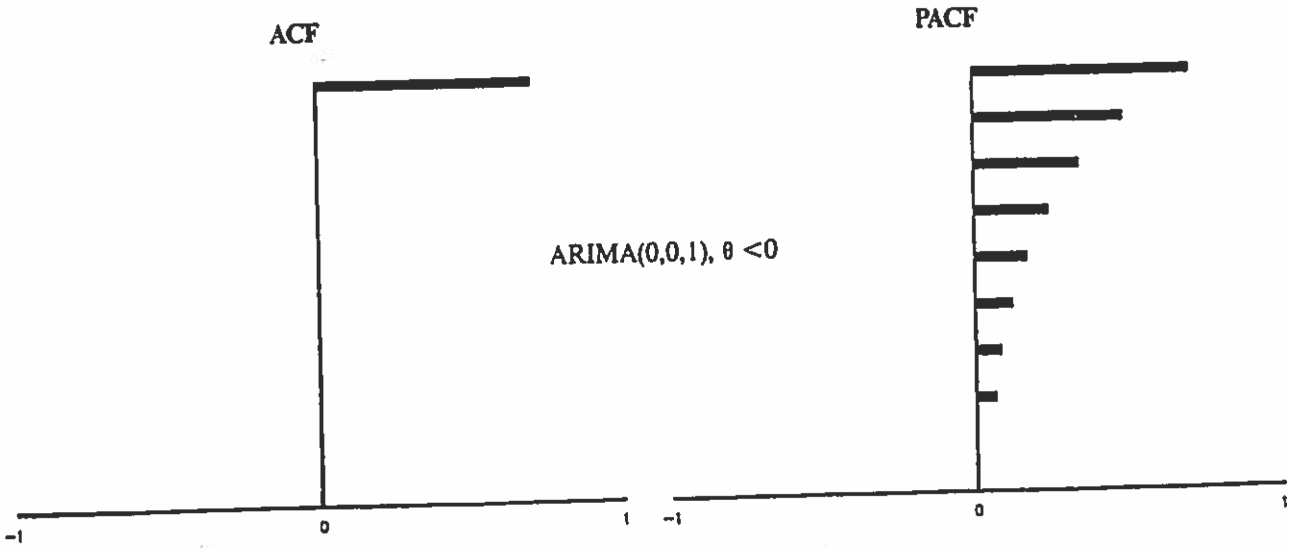


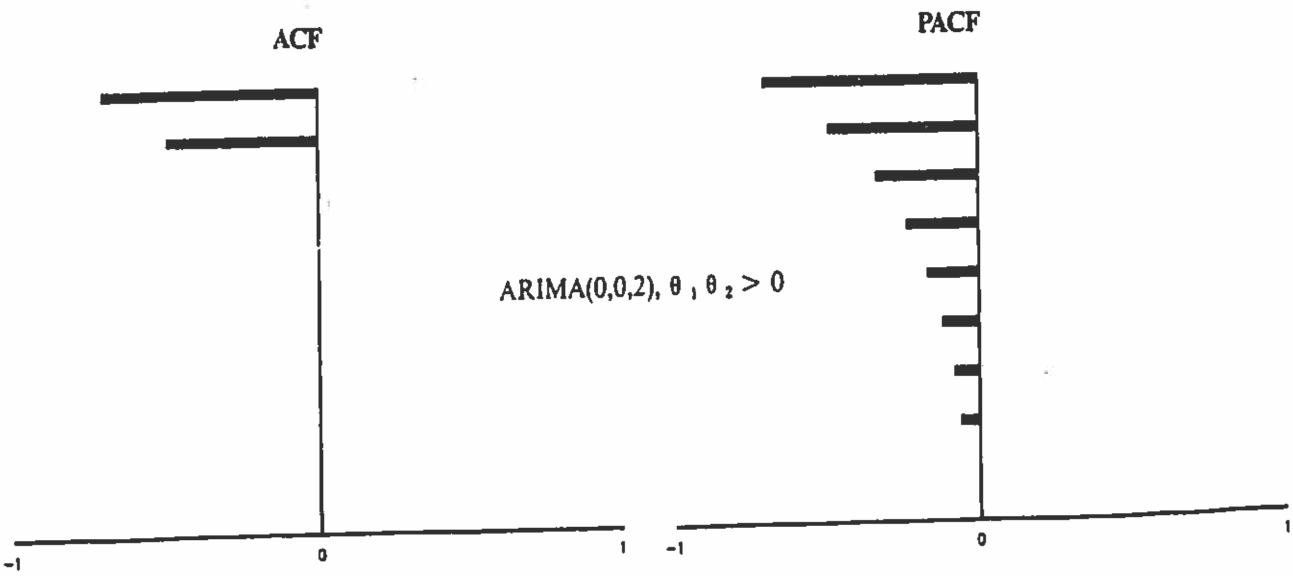


ARIMA(1,0,1), $\phi < 0$, $\theta > 0$




ACF
PACF
ARIMA(0,0,1), $\theta > 0$
-1
0
1
-1
0
1


ACF
PACF
ARIMA(0,0,1), $\theta < 0$
-1
0
1
-1
0
1


ACF
PACF
ARIMA(0,0,2), $\theta_1, \theta_2 > 0$
-1
0
1
-1
0
1

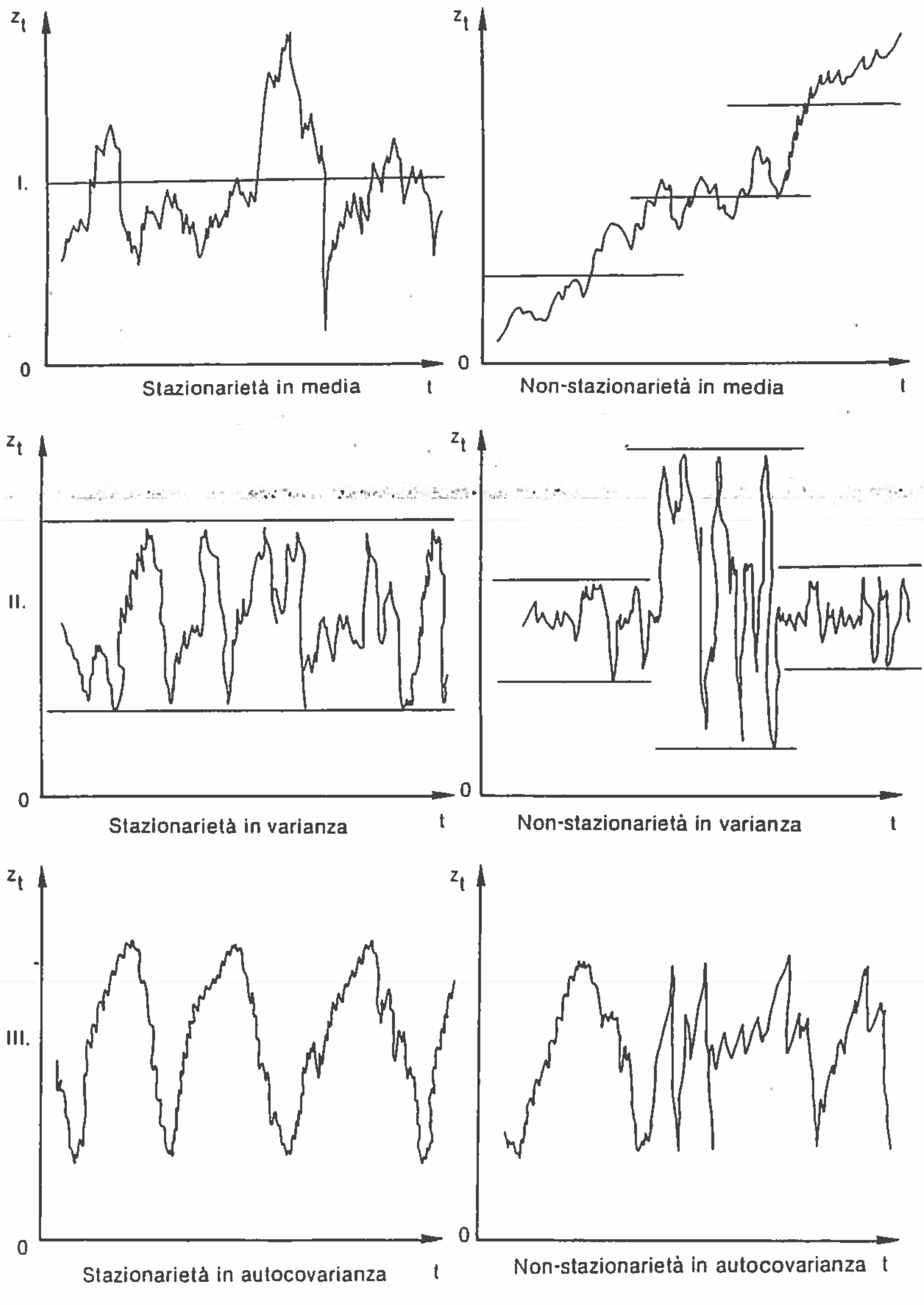


FIG. 16.5. *Caratterizzazione di serie stazionarie e serie non-stazionarie*