

# SERIE STORICHE: MODELLI LINEARI

Giovanni Fonseca  
Dipartimento di Economia  
Università degli Studi dell'Insubria - Varese

Politecnico di Torino  
20 novembre 2002

## PROCESSI STOCASTICI A TEMPO DISCRETO

$\{Y_t(\omega)\}_{t \in \mathbb{Z}}$ : un processo stocastico a tempo discreto ( $t \in \mathbb{Z}$ ) è una collezione indicizzata di variabili casuali ( $Y_t(\omega)$ ) definite su uno spazio di probabilità  $(\Omega, \mathcal{A}, P)$  a valori in  $(\mathbb{R}, \mathcal{B})$ .

$Y_{t_0}(\omega)$  (con  $t_0$  fissato) è una variabile casuale (cioè una funzione misurabile di  $\omega$ )

$Y_t(\omega_0)$  (con  $\omega_0$  fissato) è una successione di numeri reali (*realizzazione* o *traiettoria* del processo)

$$\Rightarrow Y_{t_0}(\omega_0) \in \mathbb{R}$$

## MOMENTI DI UN PROCESSO STOCASTICO

- Valore atteso

$$E[Y_t] = \mu_t$$

- Varianza

$$\text{Var}[Y_t] = E[(Y_t - \mu_t)^2] = \sigma_t^2$$

- Autocovarianza

$$\text{Cov}[Y_t, Y_{t-k}] = E[(Y_t - \mu_t)(Y_{t-k} - \mu_{t-k})] = \gamma_{t,k}$$

$$\Rightarrow \gamma_{t,0} = \sigma_t^2$$

- $Y_t = \beta t + \varepsilon_t$  con  $\varepsilon_t$  IID  $N(0, \sigma^2)$

$$\mu_t = \beta t$$

$$\begin{aligned} \gamma_{t,k} &= E[Y_t Y_{t-k}] - \beta^2 t(t-k) = \\ &E[\beta^2 t(t-k) + \varepsilon_t \beta(t-k) + \varepsilon_{t-k} \beta t + \varepsilon_t \varepsilon_{t-k}] = \end{aligned}$$

$$= \begin{cases} 0 & k > 0 \\ E[\varepsilon^2] = \sigma^2 & k = 0 \end{cases}$$

## STAZIONARIETA' DEBOLE (o del II ordine)

$$\mu_t = \mu \quad \forall t$$

$$\gamma_{t,k} = \gamma_k \quad \forall t, k \quad (\Rightarrow \sigma_t^2 = \sigma^2 \quad \forall t)$$

$$\begin{aligned} \Rightarrow \gamma_k &= \text{Cov}[Y_t, Y_{t-k}] = \text{Cov}[Y_{t+k}, Y_{[t+k]-k}] = \\ &= \text{Cov}[Y_{t+k}, Y_t] = \gamma_{-k} \end{aligned}$$

## STAZIONARIETA' FORTE (o in senso stretto)

$$\forall k, t_1, \dots, t_k$$

$$F(Y_t, Y_{t+t_1}, \dots, Y_{t+t_k}) = F(t_1, \dots, t_k)$$

- Se  $\{Y_t\}_{t \in \mathbb{Z}}$  è *fortemente* stazionario e  $E[Y_1^2] < \infty$  allora

$\{Y_t\}_{t \in \mathbb{Z}}$  è anche *debolmente* stazionario.

- Se  $\{Y_t\}_{t \in \mathbb{Z}}$  è *debolmente* stazionario e il processo è gaussiano (i.e.  $\forall k, t_1, \dots, t_k \ (Y_t, Y_{t+t_1}, \dots, Y_{t+t_k}) \sim N_{k+1}$ ) allora

$\{Y_t\}_{t \in \mathbb{Z}}$  è anche *fortemente* stazionario.

## AUTOCORRELAZIONE (globale)

Misura il legame *lineare* fra due componenti del processo.

$$\text{Corr}[Y_t, Y_{t-k}] = \rho_{t,t-k} = \frac{\gamma_{t,k}}{\sqrt{\gamma_{t,0}\gamma_{t-k,0}}}$$

Se il processo è *debolmente* stazionario allora

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

Proprietà :

- $\rho_k \in [-1, 1]$
- $\rho_0 = 1$
- $\rho_k = \rho_{-k}$

E' utile nella identificazione dell'ordine di un modello lineare.

Allo stesso scopo si calcola la funzione di *auto-correlazione parziale*.

## AUTOCORRELAZIONE PARZIALE

Misura il legame *lineare* fra due componenti  $Y_t$  e  $Y_{t-k}$  del processo *al netto* del legame lineare con le componenti intermedie  $Y_{t-1}, \dots, Y_{t-k+1}$ .

$$\text{Corr}[Y_t, Y_{t-k} | Y_{t-1}, \dots, Y_{t-k+1}] = \pi_k$$

- Come calcolare  $\pi_k$

1)

$$\pi_k = \frac{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & 1 \end{vmatrix}}$$

( $|\cdot|$  = funzione determinante della matrice)

2) Stimare la seguente *regressione lineare*

$$Y_t = a_1 Y_{t-1} + \cdots + a_{k-1} Y_{t-k+1} + \pi_k Y_{t-k} + e_t$$

## WHITE NOISE

$\{\varepsilon_t\}_{t \in \mathbb{Z}}$  è un processo  $WN(0, \sigma^2)$  se

a)  $E[\varepsilon_t] = 0 \quad \forall t$

b)  $E[\varepsilon_t^2] = \sigma^2 < \infty \quad \forall t$

c)  $E[\varepsilon_t \varepsilon_s] = 0 \quad \forall t \neq s$

- Condizioni alternative:

1)  $\varepsilon_t$  stocasticamente *indipendente* da  $\varepsilon_s$

$\forall t \neq s$

al posto di c)

2)  $\varepsilon_t \sim N(0, \sigma^2) \quad \forall t$

al posto di a), b).

- $WN$  sono le *innovazioni* o *errori* di un modello per serie storiche.



## DECOMPOSIZIONE DI WOLD

Ogni processo *debolmente* stazionario  $\{Y_t\}_{t \in \mathbb{Z}}$  può essere decomposto in due componenti di cui una deterministica e l'altra stocastica

$$Y_t = f(t) + \sum_{i=0}^{+\infty} \psi_i \varepsilon_{t-i}$$

dove  $\varepsilon_t \sim WN$ ,  $\psi_0 = 1$  e  $\sum_{i=0}^{+\infty} \psi_i^2 < +\infty$ .

## INVERTIBILITA'

Dato un processo stazionario si può calcolare univocamente la sua funzione di autocorrelazione. NON è vero il contrario, cioè ci possono essere più processi con la stessa funzione di autocorrelazione.

MA uno solo di questi processi sarà anche *invertibile*.

$\{Y_t\}_{t \in \mathbb{Z}}$  è detto *invertibile* se esiste una funzione *lineare*  $h(\cdot)$  e un processo  $WN$   $\{\varepsilon_t\}$  tali che

$$Y_t = h(Y_{t-1}, Y_{t-2}, \dots) + \varepsilon_t$$

## STIME DEI MOMENTI

Si possiede la serie storica dei dati osservati  $(y_1, \dots, y_T)$  di un processo stazionario.

- $\hat{\mu} = \frac{1}{T} \sum_{t=1}^T y_t = \bar{y}$

$\hat{\mu}$  è *non distorto*.

E' anche *consistente* se  $\sum_{i=1}^{\infty} |\rho_i| < \infty$   
(condizione di *ergodicità in media*)

N.B. Se il processo è *gaussiano* tale condizione assicura la consistenza dei momenti di ogni ordine.

- $\hat{\gamma}_k = \frac{1}{T} \sum_{t=1}^{T-k} (y_t - \hat{\mu})(y_{t+k} - \hat{\mu})$

$\hat{\gamma}_k$  è asintoticamente non distorto.

$\Downarrow$

- $\hat{\sigma}^2 = \hat{\gamma}_0 = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{\mu})^2$

- $\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$

PROPRIETA' ASINTOTICHE DI  $\hat{\rho}_k$  e  $\hat{\pi}_k$   
PER SERIE PURAMENTE CASUALI

$$E[\hat{\rho}_k] \cong -\frac{1}{T} \qquad \text{Var}[\hat{\rho}_k] \cong \frac{1}{T}$$

$$\hat{\rho}_k \xrightarrow{d} N\left(-\frac{1}{T}, \frac{1}{T}\right)$$

$$\Downarrow$$

$$IC_{0.95}(\rho_k) \cong \left(-\frac{2}{\sqrt{T}}, \frac{2}{\sqrt{T}}\right)$$

- $IC_{0.95}(\pi_k) \cong \left(-\frac{2}{\sqrt{T}}, \frac{2}{\sqrt{T}}\right)$
- Portmanteau test sui residui di un modello lineare (Box, Jenkins 1970)

$$Q = T \sum_{k=1}^M \hat{\rho}_k^2 \sim \chi_{M-p-q}^2 \qquad M \in (15, 30)$$

## OPERATORE DIFFERENZA $L$

$$LY_t = Y_{t-1}$$

$$L^k Y_t = Y_{t-k}$$

- Possiede le proprietà algebriche della moltiplicazione:
  - associativa
  - distributiva rispetto alla somma
  - commutativa rispetto alla moltiplicazione
- $Lc = c$  per  $c$  costante
- Se  $c \in (-1, 1)$  allora

$$(1 - cL)^{-1} = \sum_{i=0}^{\infty} c^i L^i$$

## MODELLI LINEARI

Nel seguito  $\{\varepsilon_t\} \sim WN(0, \sigma^2)$

- Modelli a media mobile  $MA(q)$

$$Y_t = m + \sum_{j=1}^q b_j \varepsilon_{t-j} + \varepsilon_t$$

$$Y_t = m + (1 + \sum_{j=1}^q b_j L^j) \varepsilon_t = m + b(L) \varepsilon_t$$

- Modelli autoregressivi  $AR(p)$

$$Y_t = m + \sum_{i=1}^p a_i Y_{t-i} + \varepsilon_t$$

$$a(L)Y_t = (1 - \sum_{i=1}^p a_i L^i) Y_t = m + \varepsilon_t$$

- Modelli  $ARMA(p, q)$

$$Y_t = m + \sum_{i=1}^p a_i Y_{t-i} + \sum_{j=1}^q b_j \varepsilon_{t-j} + \varepsilon_t$$

$$a(L)Y_t = m + b(L)\varepsilon_t$$

$$MA(1)$$

$$Y_t = m + b_1 \varepsilon_{t-1} + \varepsilon_t = m + (1 + b_1 L) \varepsilon_t$$

$$E[Y_t] = m$$

$$\begin{aligned} \gamma_{t,0} &= E[(Y_t - m)^2] = E[\varepsilon_t^2 + b_1^2 \varepsilon_{t-1}^2 + 2b_1 \varepsilon_t \varepsilon_{t-1}] = \\ &= (1 + b_1^2) \sigma^2 = \gamma_0 \end{aligned}$$

$$\begin{aligned} \gamma_1 &= E[(\varepsilon_t + b_1 \varepsilon_{t-1})(\varepsilon_{t-1} + b_1 \varepsilon_{t-2})] = b_1 \sigma^2 \\ \gamma_k &= 0 \quad k > 1 \end{aligned}$$

$$\rho_1 = \frac{b_1}{1+b_1^2} \quad \rho_k = 0 \quad k > 1$$

$\Rightarrow MA(1)$  debolmente stazionario

$MA(1)$  invertibile solo se  $|b_1| < 1$  perchè

$$Z_t = m + \varepsilon_t + \frac{1}{b_1} \varepsilon_{t-1}$$

possiede la stessa funzione di autocorrelazione di  $Y_t$

$$Y_t(1 + b_1 L)^{-1} = \varepsilon_t \quad \Rightarrow \quad Y_t \sum_{i=0}^{\infty} (-b_1)^i L^i = \varepsilon_t$$

$$Y_t = \varepsilon_t + \sum_{i=1}^{\infty} (-b_1)^i Y_{t-i}$$

$$MA(q)$$

$$\begin{aligned} Y_t &= m + \sum_{j=1}^q b_j \varepsilon_{t-j} + \varepsilon_t = \\ &= m + (1 + \sum_{j=1}^q b_j L^j) \varepsilon_t = m + b(L) \varepsilon_t \end{aligned}$$

$$E[Y_t] = m$$

$$\gamma_0 = \sigma^2 (1 + \sum_{j=1}^q b_j^2)$$

$$\begin{aligned} 2 \leq k \leq q \\ \gamma_k &= \sigma^2 (b_k + b_1 b_{k+1} + b_2 b_{k+2} + \cdots + b_{q-k} b_q) \end{aligned}$$

$$k > q \quad \gamma_k = 0$$

$\Rightarrow MA(q)$  debolmente stazionario.

*Invertibile* se le  $q$  radici  $z_1, \dots, z_q$  di

$$b(z) = 1 + \sum_{j=1}^q b_j z^j = 0$$

sono tali che  $|z_i| > 1 \quad i = 1, \dots, q$

- Inoltre  $\pi_k$  decadono verso 0 con  $k \rightarrow \infty$ .

$$MA(2)$$

$$\gamma_0 = \sigma^2(1 + b_1^2 + b_2^2)$$

$$\gamma_1 = \sigma^2(b_1 + b_2 b_1)$$

$$\gamma_2 = \sigma^2 b_2$$

$$MA(\infty)$$

$$Y_t = m + \varepsilon_t + \sum_{j=1}^{\infty} b_j \varepsilon_{t-j}$$

Se  $\sum_{j=1}^{\infty} |b_j| < \infty$  allora il processo è *stazionario*.



$$AR(1)$$

$$Y_t = m + a_1 Y_{t-1} + \varepsilon_t$$

$$(1 - a_1 L)Y_t = m + \varepsilon_t$$

$$\begin{aligned} \text{Se } |a_1| < 1 \text{ allora } Y_t &= \frac{m + \varepsilon_t}{1 - a_1 L} = (m + \varepsilon_t) \sum_{i=0}^{\infty} a_1^i L^i = \\ &= m \sum_{i=0}^{\infty} a_1^i + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i} \end{aligned}$$

$$E[Y_t] = \frac{m}{1 - a_1}$$

$$\gamma_0 = E\left[\left(\sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i}\right)^2\right] = \sigma^2 \sum_{i=0}^{\infty} a_1^{2i} = \frac{\sigma^2}{1 - a_1^2}$$

$$k < 0$$

$$\begin{aligned} \gamma_k &= E\left[\left(\sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i}\right)\left(\sum_{i=0}^{\infty} a_1^i \varepsilon_{t-k-i}\right)\right] = \\ &E[a_1^k \varepsilon_{t-k} + a_1^{k+1} a_1 \varepsilon_{t-k-1} + a_1^{k+2} a_1^2 \varepsilon_{t-k-2} + \dots] = \\ &= a_1^k \sigma^2 (1 + a_1^2 + a_1^4 + \dots) = \frac{a_1^k}{1 - a_1^2} \sigma^2 \end{aligned}$$

$$\Rightarrow \quad \rho_k = a_1^k$$

Quindi se  $|a_1| < 1$  il processo è *stazionario*.

$AR(1)$  : calcolo dei momenti (2)

Si assume la stazionarietà

$$\mu = E[Y_t] = E[m + a_1 Y_{t-1} + \varepsilon_t] = m + a_1 \mu$$

$$\Rightarrow \quad \mu = \frac{m}{1-a_1}$$

$$\gamma_0 = \text{Var}[Y_t] = \text{Var}[m + a_1 Y_{t-1} + \varepsilon_t] = a_1^2 \gamma_0 + \sigma^2$$

$$\Rightarrow \quad \gamma_0 = \frac{\sigma^2}{1-a_1^2} \quad \Rightarrow \quad a_1^2 < 1$$

$$\begin{aligned} \gamma_k &= E[(Y_t - \mu)(Y_{t-k} - \mu)] = \\ &= E[(\mu(1 - a_1) + a_1 Y_{t-1} + \varepsilon_t - \mu)(Y_{t-k} - \mu)] \\ &= E[(a_1(Y_{t-1} - \mu) + \varepsilon_t)(Y_{t-k} - \mu)] = \\ &= a_1 \gamma_{k-1} \end{aligned}$$

$$\Rightarrow \quad \gamma_k = a_1^k \gamma_0$$

$$AR(2)$$

$$Y_t = m + a_1 Y_{t-1} + a_2 Y_{t-2} + \varepsilon_t$$

$$(1 - a_1 L - a_2 L^2) Y_t = a(L) Y_t = m + \varepsilon_t$$

Se le due radici  $z_1$  e  $z_2$  di

$$a(z) = 1 - a_1 z - a_2 z^2 = 0$$

sono tali che  $|z_i| > 1$   $i = 1, 2$  allora il processo è *stazionario*.

$$\mu = m + a_1 \mu + a_2 \mu \quad \Rightarrow \quad \mu = \frac{m}{1 - a_1 - a_2}$$

$$\Rightarrow Y_t = \mu(1 - a_1 - a_2) + a_1 Y_{t-1} + a_2 Y_{t-2} + \varepsilon_t$$

$$\Rightarrow Y_t - \mu = a_1(Y_{t-1} - \mu) + a_2(Y_{t-2} - \mu) + \varepsilon_t$$

Allora per  $k > 1$

$$\gamma_k = a_1 \gamma_{k-1} + a_2 \gamma_{k-2}$$

$$\rho_k = a_1 \rho_{k-1} + a_2 \rho_{k-2}$$

$$\rho_1 = a_1 + a_2 \rho_1 \quad \Rightarrow \quad \rho_1 = \frac{a_1}{1 - a_2}$$

$$\rho_2 = a_1 \rho_1 + a_2$$

$$\gamma_0 = a_1 \gamma_1 + a_2 \gamma_2 + \sigma^2 = a_1 \gamma_0 \rho_1 + a_2 \gamma_0 \rho_2 + \sigma^2$$

$$\Rightarrow \gamma_0 = \frac{(1 - a_2) \sigma^2}{(1 + a_2)[(1 - a_2^2) - a_1^2]}$$

$$AR(p)$$

$$Y_t = m + \sum_{i=1}^p a_i Y_{t-i} + \varepsilon_t$$

$$(1 - \sum_{i=1}^p a_i L^i) Y_t = a(L) Y_t = m + \varepsilon_t$$

Se le  $p$  radici  $z_1, \dots, z_p$  di  $a(z) = 0$  sono tali che  $|z_i| > 1$   $i = 1, \dots, p$  allora il processo è *stazionario*.

Ovviamente  $\pi_k = 0$  per  $k > p$ .

$$\mu = \frac{m}{1 - a_1 - \dots - a_p}$$

$$\gamma_k = \begin{cases} a_1 \gamma_1 + \dots + a_p \gamma_p + \sigma^2 & k = 0 \\ a_1 \gamma_{k-1} + \dots + a_p \gamma_{k-p} & k > 0 \end{cases}$$

Si risolve il sistema lineare di  $p + 1$  equazioni nelle incognite  $\gamma_0, \dots, \gamma_p$  in funzione di  $a_1, \dots, a_p, \sigma^2$ .

• Sistema di equazioni di *Yule-Walker*

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \dots & \dots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-2} \\ \vdots & & \ddots & & \vdots \\ \rho_{p-1} & \rho_{p-2} & \dots & \rho_1 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}$$

$$ARMA(p, q)$$

$$Y_t = m + \sum_{i=1}^p a_i Y_{t-i} + \sum_{j=1}^q b_j \varepsilon_{t-j} + \varepsilon_t$$

$$a(L)Y_t = m + b(L)\varepsilon_t$$

$$\mu = \frac{m}{1 - a_1 - \dots - a_p}$$

- Se le  $p$  radici  $z_1, \dots, z_p$  di

$$a(z) = 1 - \sum_{i=1}^p a_i z^i = 0$$

sono tali che

$|z_i| > 1$   $i = 1, \dots, p$  allora il processo è *stazionario*.

- Se le  $q$  radici  $z_1, \dots, z_q$  di

$$b(z) = 1 + \sum_{j=1}^q b_j z^j = 0$$

sono tali che

$|z_i| > 1, i = 1, \dots, q$  allora il processo è *invertibile*.

- Per  $k > q$  le autocovarianze (autocorrelazioni) hanno la stessa struttura di quelle relative a un processo  $AR(p)$

$$\rho_k = a_1 \rho_{k-1} + \dots + a_p \rho_{k-p}$$

- Per  $k > p$  le autocorrelazioni parziali  $\pi_k$  hanno la stessa struttura di quelle relative a un processo  $MA(q)$

$$ARMA(p, q) \quad (2)$$

Un modello  $ARMA(p, q)$  si assume sempre essere in forma *ridotta*, cioè si assume che le  $p$  radici  $\lambda_i$ ,  $i = 1, \dots, p$  di  $a(z) = 0$  siano tutte diverse dalle  $q$  radici  $\eta_j$ ,  $j = 1, \dots, q$  di  $b(z) = 0$ .

Altrimenti, se esiste  $\lambda_i = \eta_j$  per qualche  $i, j$ , si ha che nella fattorizzazione

$$\begin{aligned} (1 - \lambda_1 L) \cdots (1 - \lambda_p L)(Y_t - \mu) &= \\ &= (1 - \eta_1 L) \cdots (1 - \eta_q L) \varepsilon_t \end{aligned}$$

si possono semplificare i due fattori uguali.

Il modello quindi si *riduce* al modello  $ARMA(p - 1, q - 1)$ .

## IDENTIFICAZIONE

Si assume che il modello generatore dei dati sia *stazionario*.

In questo caso dall'analisi dei grafici delle stime delle autocorrelazioni, globali e parziali, ci si può fare un'idea del modello generatore dei dati osservati.

- Se le autocorrelazioni globali  $\hat{\rho}_k$  si annullano da un certo istante temporale  $q$  e le autocorrelazioni parziali  $\hat{\pi}_k$  decadono verso 0 esponenzialmente od oscillando col crescere di  $k$  allora il modello candidato è  $MA(q)$ .

$$(\hat{\rho}_k = 0 \quad k > q \quad \hat{\pi}_k \rightarrow 0)$$

- Analogamente se le autocorrelazioni parziali  $\hat{\pi}_k$  si annullano da un certo istante temporale  $p$  e le autocorrelazioni globali  $\hat{\rho}_k$  decadono verso 0 esponenzialmente od oscillando col crescere di  $k$  allora il modello candidato è  $AR(p)$ .

$$(\hat{\pi}_k = 0 \quad k > p \quad \hat{\rho}_k \rightarrow 0)$$

## IDENTIFICAZIONE (2)

- Se si ricade in una situazione diversa dalle due precedenti allora si usa un modello  $ARMA(p, q)$ . Per determinare gli ordini  $p, q$  si possono usare diversi criteri fra cui  $AIC$  (Akaike, 1974) dato da

$$AIC(k) \cong (T - p) \ln \hat{\sigma}^2 + 2k$$

con  $k = p + q + 2$  il numero di parametri stimati,  $\hat{\sigma}^2$  la stima della varianza delle innovazioni che rappresenta anche la varianza residua del modello stimato.

- In caso di NON *stazionarietà* , cioè quando:
  - 1) le autocorrelazioni globali non decadono verso 0, e.g. presenza di trend
  - 2) dal grafico della serie dei dati osservati si nota eteroschedasticità , e.g. varianza che cresce col tempo
  - 3) presenza di stagionalità

bisogna trasformare la serie al fine di ottenere una serie stazionaria.



## TRASFORMAZIONI DEI DATI

1) Differenziazioni successive dei dati finchè non si ottiene una serie stazionaria  $\Rightarrow$   $ARIMA(p, d, q)$  processo *integrato* di ordine  $d$

$$Y_t \longrightarrow (1 - L)^d Y_t$$

2) Trasformazioni di Box-Cox (1964)  
(usata anche per rendere i dati gaussiani)

$$T(Y_t) = \begin{cases} \frac{Y_t^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln Y_t & \lambda = 0 \end{cases}$$

Di solito si applicano le trasformazioni con  $\lambda = 0, \frac{1}{2}$ .

3) Destagionalizzazione tramite medie mobili oppure incorporando le componenti stagionali nel modello  $\Rightarrow SARMA(p, q) \times (P, Q)_s$

$$a(L)A(L^s)Y_t = m + b(L)B(L^s)\varepsilon_t$$

con  $A(L^s) = 1 - A_1 L^s - A_2 L^{2s} - \dots - A_P L^{Ps}$  e  
 $B(L^s) = 1 + B_1 L^s + B_2 L^{2s} + \dots + B_Q L^{Qs}$

## STIME DI MASSIMA VEROSIMIGLIANZA CONDIZIONATA (ai valori iniziali)

- Si assume che il processo delle innovazioni sia *gaussiano* ( $\varepsilon_t \text{ IID } N(0, \sigma^2)$ ).

In ogni caso le stime sono *consistenti* anche senza l'ipotesi di gaussianità .

$y_1, \dots, y_T$  dati osservati.

- Per costruire la verosimiglianza si usano le densità condizionate ai valori passati della serie. Quindi i primi dati della serie osservata saranno usati come valori iniziali *fissi*. I valori iniziali delle innovazioni sono posti uguali a 0 (pari al loro valore atteso).

- Di solito la procedura di massimizzazione è di tipo numerico (grid search, Newton-Raphson, etc.) e necessita di valori di partenza per i parametri da stimare.

Si possono usare come valori di partenza le stime dei parametri ottenute dalle equazioni di *Yule-Walker* con i parametri del modello come incognite e le autocorrelazioni empiriche note.

- Software statistici (e.g. E-views) calcolano le stime di massima verosimiglianza esatta.

## STIMA $AR(p)$

$$\begin{aligned} L(\theta) &= f_{Y_T, Y_{T-1}, \dots, Y_{p+1} | Y_1, \dots, Y_p} = \\ &= \prod_{t=p+1}^T f_{Y_t | Y_{t-1}, \dots, Y_{t-p}} \end{aligned}$$

$$Y_i = y_i \quad i = 1, \dots, p$$

$$Y_t | Y_{t-1}, \dots, Y_{t-p} \sim N(m + a_1 Y_{t-1} + \dots + a_p Y_{t-p}, \sigma^2)$$

$$\hat{\sigma}^2 = \frac{1}{T-p} \sum_{t=p+1}^T \hat{\varepsilon}_t^2$$

$$\text{con } \hat{\varepsilon}_t = y_t - \hat{a}_1 y_{t-1} \quad t = p+1, \dots, T.$$

- Le stime sono le stesse che si ottengono da una procedura OLS.

- $AR(1)$

$$L(\theta) = \prod_{t=2}^T f_{Y_t | Y_{t-1}}$$

$$Y_1 = y_1$$

$$Y_t | Y_{t-1} \sim N(m + a_1 Y_{t-1}, \sigma^2)$$

- Si ottengono stime consistenti anche se  $|a_1| > 1$

$$MA(q)$$

- Si assume l'invertibilità del modello.

$$\text{Valori iniziali } \varepsilon_0 = \varepsilon_{-1} = \dots = \varepsilon_{-q+1} = 0$$

$$\Downarrow$$

$$\varepsilon_1 = y_1 - m \quad \varepsilon_2 = y_2 - m - b_1(y_1 - m) \quad \dots$$

$$\Downarrow$$

$$L(\theta) = f_{Y_1|\underline{\varepsilon}_0} \prod_{t=2}^T f_{Y_t|Y_{t-1}, \dots, Y_1, \underline{\varepsilon}_0}$$

$$Y_t|Y_{t-1}, \dots, Y_1, \underline{\varepsilon}_0 \sim N(m + \sum_{j=1}^q b_j \varepsilon_{t-j}, \sigma^2)$$

$$ARMA(p, q)$$

Valori iniziali

$$Y_i = y_i \quad i = 1, \dots, p$$

$$\varepsilon_0 = \varepsilon_{-1} = \dots = \varepsilon_{-q+1} = 0$$

$$L(\theta) = \prod_{t=p+1}^T f_{Y_t|Y_{t-1}, \dots, Y_1, \underline{\varepsilon}_0}$$

## PREVISIONE (puntuale)

Sia  $I_T = \{Y_T, Y_{T-1}, \dots, Y_0, Y_{-1}, \dots\}$  l'insieme di *informazione passata* su un processo  $\{Y_t\}_{t \in \mathbb{Z}}$  fino al tempo  $T$ .

SCOPO: Prevedere il valore *futuro* di  $Y_{T+k}$ , avendo come informazioni  $I_T$ , tramite un previsore  $\hat{Y}_{T+k|T}$  con buone proprietà .

In particolare, si ricerca il *previsore ottimo*, cioè il previsore che minimizza l'errore quadratico medio della previsione

$$\text{EQM}(\hat{Y}_{T+k|T}) = E[(Y_{T+k} - \hat{Y}_{T+k|T})^2].$$

$$\Rightarrow \hat{Y}_{T+k|T} = E[Y_{T+k}|I_T]$$

In realtà non si possiede  $I_T$  e il *vero* modello generatore della serie, ma solo un modello *stimato* tramite la serie dei dati osservati  $\{y_t\}_{t=1}^T$ .

## PREVISIONE (puntuale) (2)

- $ARMA(p, q)$

$$\begin{aligned}\hat{Y}_{T+k|T} &= E[Y_{T+k}|I_T] = \\ &= E[m + \sum_{i=1}^p a_i Y_{T+k-i} + \sum_{j=1}^q b_j \varepsilon_{T+k-j} + \varepsilon_{T+k}|I_T]\end{aligned}$$

$$E[Y_{T+j}|Y_T = y_T, \dots, Y_1 = y_1] = \begin{cases} y_{T+j} & j \leq 0 \\ \hat{y}_{T+j|T} & j > 0 \end{cases}$$

$$E[\varepsilon_{T+j}|Y_T = y_T \dots Y_1 = y_1] = \begin{cases} \hat{\varepsilon}_{T+j} = Y_{T+j} - \hat{Y}_{T+j|T+j-1} & j \leq 0 \\ 0 & j > 0 \end{cases}$$

## PREVISIONE (puntuale): Esempi

- N.B. Se si è trasformata la serie per renderla stazionaria o gaussiana le previsioni dovranno subire la trasformazione inversa per riferirsi al fenomeno originario.

- $AR(1)$

$$\hat{y}_{T+1} = m + a_1 y_T$$

$$\hat{y}_{T+2} = m + a_1 \hat{y}_{T+1} \quad \dots$$

- $MA(1)$

$$\varepsilon_1 = y_1 - m$$

$$\Rightarrow \hat{y}_{T+1} = m + b_1(y_T - m - b_1 \varepsilon_{T-1}) = \dots$$

$$\hat{y}_{T+2} = m$$

$$\hat{y}_{T+k} = m \quad k > 2$$

- $ARMA(1, 1)$

$$\varepsilon_1 = y_1 - m$$

$$\Rightarrow \hat{y}_{T+1} = m + a_1 y_T + b_1(y_T - a_1 y_{T-1} - m - b_1 \varepsilon_{T-1}) = \dots$$

$$\hat{y}_{T+2} = m + a_1 \hat{y}_{T+1}$$

$$\hat{y}_{T+k} = m + a_1 \hat{y}_{T+k-1} \quad k > 1$$

## PREVISIONE (intervallare)

SCOPO: Determinare  $x_l, x_u$  tali che  
 $\Pr[x_l \leq Y_{T+k} \leq x_u] = 1 - \alpha$  con  $\alpha$  prefissato.

- Procedura *approssimata*. Ipotesi:
  - 1) il processo delle innovazioni è *gaussiano*
  - 2) il modello è correttamente identificato
  - 3) i parametri sono *noti*
  - 4)  $I_T$  è disponibile

Definiamo l'*errore di previsione*

$$e_{T+k} = Y_{T+k} - \hat{Y}_{T+k|T}$$

$e_{T+k}$  è una variabile aleatoria, funzione lineare delle innovazioni future rispetto al tempo  $T$  (e quindi ignote)  $\Rightarrow E[e_{T+k}] = 0$ .

Per ottenere l'intervallo di previsione serve la varianza dell'errore di previsione.

$$x_l = \hat{Y}_{T+k|T} - z_{\alpha/2} \sqrt{\text{Var}[e_{T+k}]}$$

$$x_u = \hat{Y}_{T+k|T} + z_{\alpha/2} \sqrt{\text{Var}[e_{T+k}]}$$

con  $z_{\alpha/2}$  il quantile di livello  $\alpha/2$  della distribuzione normale standardizzata.



## PREVISIONE (intervallare) (2)

- $AR(1)$

$$\text{Var}[e_{T+1}] = \text{Var}[\varepsilon_{T+1}] = \sigma^2$$

$$\text{Var}[e_{T+2}] = \text{Var}[\varepsilon_{T+2} + a_1\varepsilon_{T+1}] = \sigma^2(1 + a_1^2)$$

$$\begin{aligned}\text{Var}[e_{T+k}] &= \text{Var}[\varepsilon_{T+k} + a_1\varepsilon_{T+k-1} + \cdots + a_1^{k-1}\varepsilon_{T+1}] \\ &= \sigma^2(1 + a_1^2 + a_1^4 + \cdots + a_1^{2(k-1)})\end{aligned}$$

- $MA(1)$

$$\text{Var}[e_{T+1}] = \text{Var}[\varepsilon_{T+1}] = \sigma^2$$

$$\text{Var}[e_{T+2}] = \text{Var}[\varepsilon_{T+2} + b_1\varepsilon_{T+1}] = \sigma^2(1 + b_1^2)$$

$$\begin{aligned}\text{Var}[e_{T+k}] &= \text{Var}[\varepsilon_{T+k} + b_1\varepsilon_{T+k-1}] = \\ &= \sigma^2(1 + b_1^2) \quad k > 1\end{aligned}$$

- $ARMA(1, 1)$

$$\text{Var}[e_{T+1}] = \text{Var}[\varepsilon_{T+1}] = \sigma^2$$

$$\begin{aligned}\text{Var}[e_{T+2}] &= \text{Var}[a_1\varepsilon_{T+1} + \varepsilon_{T+2} + b_1\varepsilon_{T+1}] = \\ &= \sigma^2(1 + (a_1 + b_1)^2)\end{aligned}$$

## PREDIZIONE (intervallare)

E' possibile ottenere intervalli di previsione per processi  $AR$  con una probabilità di copertura migliore (soprattutto nel caso di serie di dati osservati di lunghezza limitata) facendo ricorso alla teoria asintotica della verosimiglianza.

- Questo metodo funziona anche quando, come di solito accade, le ipotesi poste in precedenza non sono tutte verificate (in particolare i parametri non sono noti, ma stimati).
- Difetti:
  - 1) difficoltà di computazione
  - 2) work in progress
- Si ottengono intervalli della distribuzione di  $Y_{T+1|T}$  con una probabilità di copertura esatta a meno di termini di ordine  $O(n^{-3/2})$ .

$AR(1)$  con parametri tutti ignoti

$$\left(\frac{\nu}{T-\tau}\right)^{\frac{1}{2}}\left[\frac{Y_{T+1}-m-\frac{(1+a_1)y_{T-m}}{T}-\left(\frac{T+1}{T}\right)a_1(y_T-m)}{\sigma}\right]\sim t_\nu$$

con

$$\nu=\frac{(T-\tau)^2}{T-1}\qquad \tau=2+\frac{(1-a_1^2)(y_T-m)^2}{\sigma^2}$$