Serie Storiche
La procedura di Box-Jenkins

Corso di Laurea in Statistica per le Tecnologie e le Scienze
Corso di Laurea in Statistica per l'Economia e l'Impresa
matricole dispari

Francesco Lisi

a.a. 2026/2027

Il problema

- Come passare dalla serie storica osservata al modello, cioè alla stima del processo stocastico che ha generato i dati?
- Se ci muoviamo dentro la classe ARIMA, la prima cosa da fare è identificare il modello, ovvero identificare gli ordini d, p, e q
  1. d, è il numero di differenziazioni che rendono stazionaria y_t.
     $$d \Rightarrow (1-B)^d y_t = x_t, \text{ con } x_t \text{ stazionaria.}$$
  2. p è l'ordine della componente autoregressiva.
  3. q è l'ordine della componente a media mobile.
- Il modello per y_t è quindi:

$$(1-B)^d y_t = x_t$$

$$x_t = \phi_0 + \sum_{i=1}^{p} \phi_i x_{t-i} + \epsilon_t - \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$$

o anche:

$$\phi(B)(1-B)^d y_t = \phi_0 + \theta(B)\epsilon_t$$

Il problema

- Una volta identificato il modello bisogna stimarne i $p+q+2$ parametri
  $$\phi_0, \phi_1, ..., \phi_p, \theta_1, ..., \theta_q, \sigma_\varepsilon^2$$
- Si noti che i parametri del modello per y_t sono gli stessi di quelli del modello per x_t. Essendo quest'ultimo stazionario li possiamo stimare sulle x_t

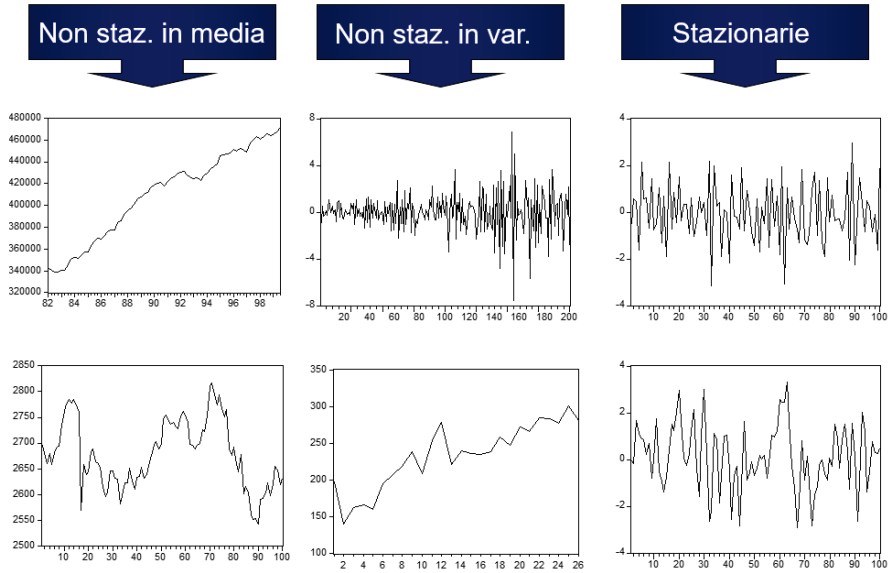
La procedura di Box e Jenkins

- La procedura di Box e Jenkins (1976) serve a costruire, a partire da una serie storica osservata, il modello ARIMA che meglio approssima il processo generatore dei dati osservati.
- Si ipotizza che la serie sia già stata resa stazionaria attraverso opportune trasformazioni preliminari (se necessarie).
- I passi della procedura Box-Jenkins sono:
  1. **Identificazione**
  2. **Stima dei parametri**
  3. **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/o previsivi.

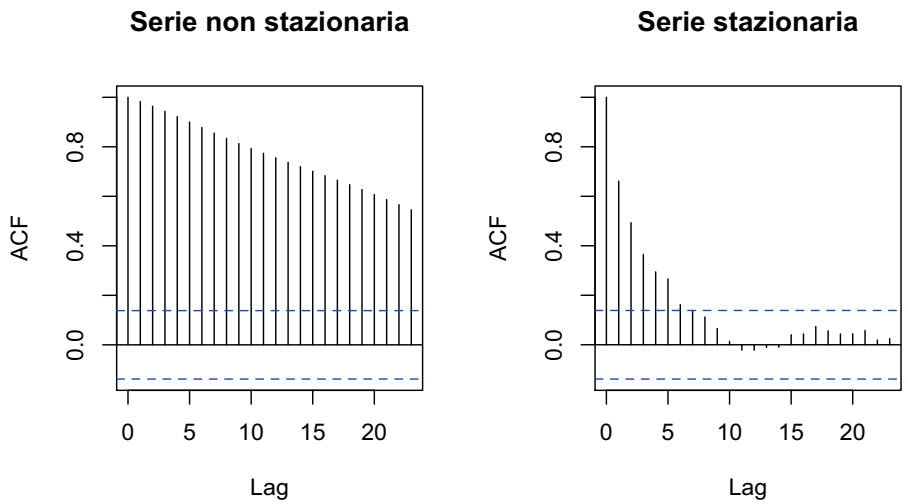
Analisi preliminari

- Sono presenti valori anomali (*outliers*)?
- La serie è stazionaria?
  a. grafico della serie
  b. ACF
- La serie è stazionaria?
  - NO $\Rightarrow$ Trasformazioni
  - SI $\Rightarrow$ Identificazione

Stazionarietà: grafico della serie



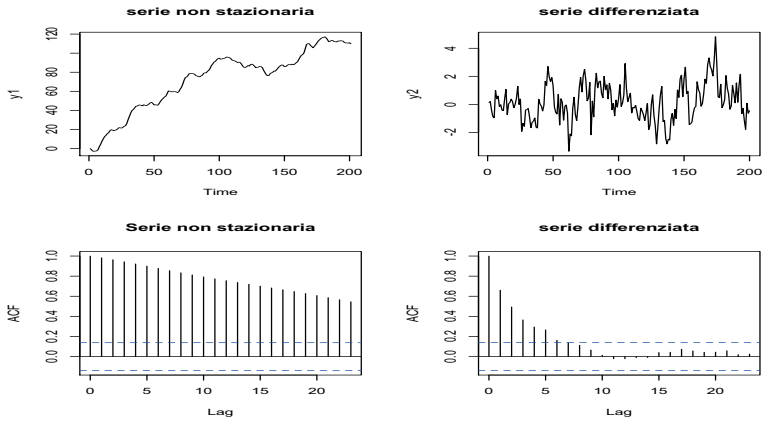
Stazionarietà: ACF



Trasformazioni che rendono stazionaria la serie

- Non stazionarietà in media: differenziazione di ordine d :

$$d = 1 \Rightarrow x_t = (1 - B)y_t = y_t - y_{t-1}$$

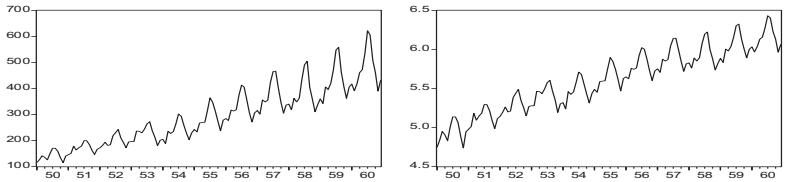


Trasformazioni che rendono stazionaria la serie

- Non stazionarietà in varianza: trasformata di Box-Cox:

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

- Come scegliere λ?
  - mediante ispezione grafica, controllando se la serie appare stabilizzata nella varianza;
  - modellando $Y_t^{(\lambda)}$ con λ ulteriore parametro da stimare.



Identificazione del modello

- Specificazione dell'ordine del modello (ordini p e q)
- d deriva dalle analisi fatte precedentemente
- Principali strumenti da usare:
  - Identificazione *strutturale* (ACF e PACF empiriche)
  - Metodi automatici di identificazione
  - Criterio della sovraparametrizzazione (solo per modelli annidati)

Identificazione strutturale

- Si parte dall'osservazione delle funzioni di autocorrelazione empiriche, tenendo in considerazione il fatto che quando si lavora con dati reali i comportamenti sono meno netti che in teoria
- Si cerca di riconoscere nelle funzioni di autocorrelazione empiriche la struttura di un modello ARMA noto.
- Più difficile nel caso di modelli misti.

Significatività delle autocorrelazioni

- Siamo interessati ad effettuare dei test di significatività del tipo:

$$\begin{cases} H_0 : \rho_k = 0 \\ H_1 : \rho_k \neq 0 \end{cases} \qquad \begin{cases} H_0 : P_k = 0 \\ H_1 : P_k \neq 0 \end{cases}$$

- Si sfrutta il fatto che sotto l'ipotesi nulla (e di gaussianità)

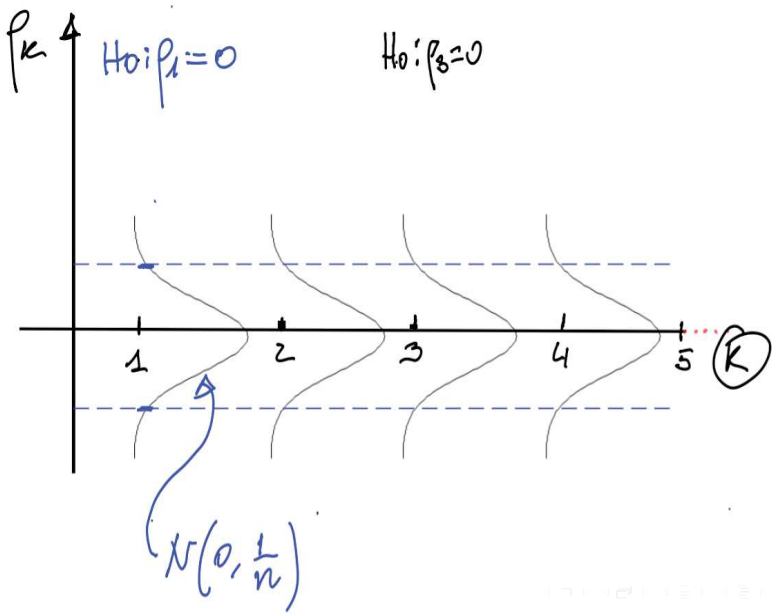
$$\hat{\rho}_k \sim N\left(0, \frac{1}{n}\right) \qquad \hat{P}_k \sim N\left(0, \frac{1}{n}\right)$$

- Si accetta H_0 se

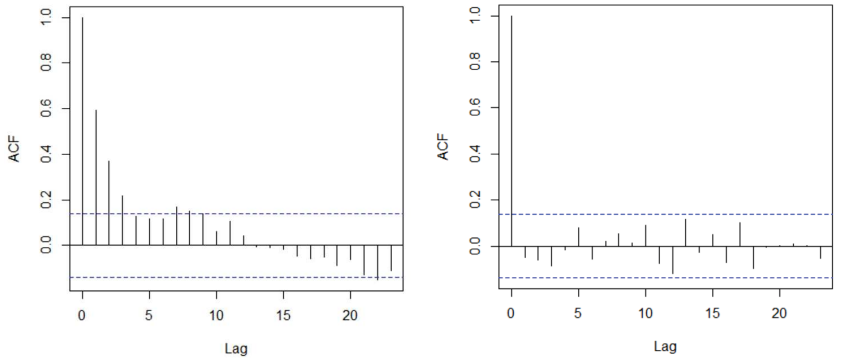
$$-z_{(1-\alpha/2)} \cdot \sqrt{Var(\hat{\rho}_k)} \leq \hat{\rho}_k \leq z_{(1-\alpha/2)} \cdot \sqrt{Var(\hat{\rho}_k)}$$

$$\frac{-1.96}{\sqrt{n}} \leq \hat{\rho}_k \leq \frac{1.96}{\sqrt{n}}$$

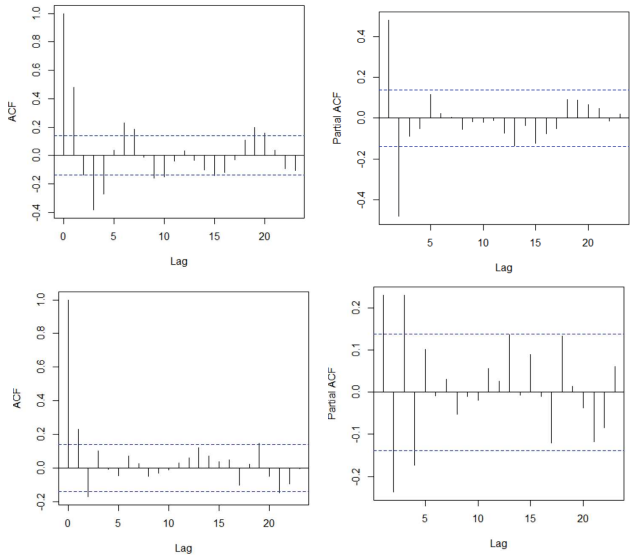
Il test di Bartlett



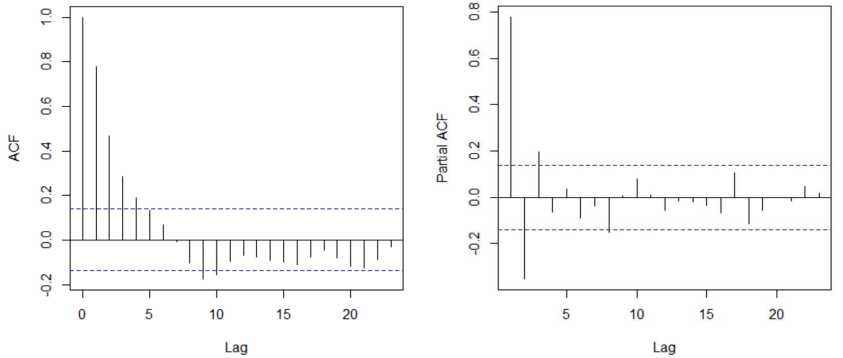
Il test di Bartlett in R



Identificazione



Identificazione



Metodi automatici di identificazione

- Se ipotizziamo per y_t un modello $ARMA(p, q)$ dobbiamo stimare i $k = p + q + 2$ parametri contenuti nel vettore $\hat{\delta} = (\phi_0, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \sigma^2_\epsilon)$ che portano al valore $logL(\hat{\delta})$ della log-verosimiglianza.
- I metodi automatici di identificazione mettono insieme due elementi:
  - [a.] la log-verosimiglianza, $logL(\hat{\delta})$, che dice quanto bene il modello si adatta ai dati,
  - [b.] un termine aggiuntivo che penalizza la complessità del modello. Quest'ultima è rappresentata dal numero di parametri presenti nel modello.
- I diversi criteri si differenziano principalmente per il peso che viene dato al termine di penalità.

Criteri automatici di identificazione

Ci sono molti criteri automatici di identificazione, tra questi:

1. Il criterio di informazione automatica di Akaike (1974):

$$AIC(k) = -\frac{2}{n} logL(\hat{\delta}) + \frac{2}{n} k$$

2. Il criterio bayesiano di Schwartz (1978):

$$BIC(k) = -\frac{2}{n} logL(\hat{\delta}) + \frac{\log n}{n} k$$

3. Il criterio di Hannan-Quinn (1979):

$$HQC(k) = -\frac{2}{n} logL(\hat{\delta}) + 2k \frac{\log \log n}{n}$$

Criteri automatici di identificazione

- Confronto fra modelli: tra due (o più) modelli si preferisce quello che ha un valore di $AIC(k)$ o $BIC(k)$ o $HQC(k)$ inferiore
- Identificazione automatica: si calcola $AIC(k)$ o $BIC(k)$ o $HQC(k)$ per vari valori di k (ossia per diversi modelli) e si scelgono gli ordini p e/o q che minimizzano il criterio scelto
- per esempio, se scegliamo l'AIC:

$$\begin{aligned} ARMA(1,0) &\rightarrow AIC(1+0+2) \\ ARMA(0,1) &\rightarrow AIC(0+1+2) \\ ARMA(1,1) &\rightarrow AIC(1+1+2) \\ ARMA(2,1) &\rightarrow AIC(2+1+2) \\ \ldots &\rightarrow \ldots \end{aligned}$$

Stima dei parametri

- Dopo aver identificato un possibile modello, il passo successivo è quello di stimarne i parametri.
- Consideriamo un modello $ARMA(p, q)$. Si tratta di stimare i parametri

$$(\phi_0, \phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q, \sigma^2_\epsilon)$$

- Ci sono diversi metodi per stimare i parametri di un modello per serie storiche.

Il metodo dei momenti (Yule-Walker)

- Il metodo dei momenti consiste nell'uguagliare un certo numero di momenti campionari teorici con i corrispondenti momenti campionari (p.e. la media campionaria $\bar{y}$, la varianza campionaria $\hat{\gamma}_0$ ecc.) e risolvere il sistema di equazioni sottostante rispetto ai parametri del modello.
- Se il modello è un $AR(p)$ è possibile utilizzare le equazioni di Yule-Walker per stimare i parametri $\phi_1, \phi_2, \ldots, \phi_p$
- Si parte dalle equazioni teoriche:

$$\begin{aligned}
\rho_1 &= \phi_1 + \phi_2\rho_1 + \phi_3\rho_2 + \ldots + \phi_p\rho_{p-1} \\
\rho_2 &= \phi_1\rho_1 + \phi_2 + \phi_3\rho_1 + \ldots + \phi_p\rho_{p-2} \\
&\vdots \\
\rho_p &= \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \phi_3\rho_{p-3} + \ldots + \phi_p
\end{aligned}$$

Il metodo dei momenti (Yule-Walker)

- Si sostituiscono i ρ_k teorici con i coefficienti empirici, $\hat{\rho}_k$, e si risolve il sistema per trovare $\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p$.
- Si usano quindi i $\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p$ e il risultato

$$\gamma_0 = \phi_1\gamma_1 + \phi_2\gamma_2 + \ldots, +\phi_p\gamma_p + \sigma_\epsilon^2$$

per ottenere uno stimatore di σ_ϵ^2 cioè:

$$\hat{\sigma}_\epsilon^2 = \hat{\gamma}_0(1 - \hat{\phi}_1\hat{\rho}_1 - \hat{\phi}_2\hat{\rho}_2 - \ldots - \hat{\phi}_p\hat{\rho}_p)$$

- La stima di μ si ottiene semplicemente sostituendola con la media campionaria, $\hat{\mu} = \bar{y}$

Minimi quadrati

- I modelli $ARMA(p, q)$ sono particolari modelli di regressione in cui le variabili esplicative sono variabili casuali.
- Gli stimatori ai minimi quadrati ordinari possono ancora essere applicati purché il termine d'errore rimanga incorrelato con le variabili esplicative, in caso contrario gli stimatori perdono la consistenza.
- In generale, per modelli $ARMA(p, q)$ questa condizione non è rispettata tranne quando $q = 0$ (modelli AR puri).

Massima verosimiglianza

- Esistono diversi metodi di stima di massima verosimiglianza che possono essere utilizzati nell'ambito delle serie storiche:
  1. Stima di MV condizionata
  2. Stima di MV non condizionata e metodo del *backcasting*
  3. Stima di MV esatta
- I primi due metodi sono approssimati.

Stima di MV esatta per un AR(1)

- Illustriamo la derivazione della MV per un modello AR(1):

$$Y_t = \phi_0 + \phi Y_{t-1} + \epsilon_t$$

  on $|\phi| < 1$ e $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$
- In tal caso i parametri da stimare sono ϕ_0, ϕ_1 e σ_ϵ^2
- Se ϵ_t è gaussiano si ha che

$$Y_t \sim \mathcal{N}\left(\frac{\phi_0}{1-\phi_1}, \frac{\sigma_\epsilon^2}{1-\phi_1^2}\right)$$

  con le Y_t non indipendenti ma autocorrelate.
- Se invece consideriamo la distribuzione di Y_t condizionatamente all'informazione fino all'istante $(t-1)$, I_{t-1}, si ha che

$$Y_t \mid I_{t-1} \sim \mathcal{N}\left(\phi_0 + \phi_1 Y_{t-1}, \sigma_\epsilon^2\right)$$

Stima di MV esatta per un AR(1)

- La verosimiglianza nel nostro caso è data da

$$L(\phi_0, \phi_1, \sigma_\epsilon^2 \mid y_1, ..., y_n) = f(y_1, ..., y_n \mid \phi_0, \phi_1, \sigma_\epsilon^2)$$

- Ora le y_t non sono indipendenti ma, posto $\delta = (\phi_0, \phi_1, \sigma_\epsilon^2)$, possimo fattorizzare la densità congiunta come

$$\begin{aligned} f(y_1, ..., y_n \mid \delta) &= f(y_n \mid I_{n-1}, \delta) \cdot f(y_{n-1} \mid I_{n-2}, \delta) \cdot ... \\ &\quad ... \cdot f(y_2 \mid I_1, \delta) \cdot f(y_1 \mid \delta) \end{aligned}$$

  con $I_t = y_t, y_{t-1}, ..., y_1$ informazione fino al tempo t.
- Dunque la verosimiglianza si può scrivere come

$$L(\delta \mid y_1, ..., y_n) = f(y_1 \mid \delta) \prod_{t=2}^{n} f(y_t \mid I_{t-1}, \delta)$$

Stima di MV esatta per un AR(1)

- E la log-verosimiglianza diventa

$$\log L(\delta \mid y_1, ..., y_n) = \log f(y_1 \mid \delta) + \sum_{t=2}^{n} \log f(y_t \mid I_{t-1}, \delta)$$

- e, tenendo conto di come sono fatte le distribuzioni condizionate e non condizionate:

$$\begin{aligned} &\log L(\delta \mid y_1, ..., y_n) = \\ &-\tfrac{1}{2}\log(2\pi) - \tfrac{1}{2}\log\left(\tfrac{\sigma_\epsilon^2}{1-\phi_1^2}\right) - \tfrac{(y_1 - (\phi_0/(1-\phi_1)))^2}{2\sigma_\epsilon^2/(1-\phi_1^2)} + \\ &-\tfrac{n-1}{2}\log(2\pi) - \tfrac{n-1}{2}\log\left(\sigma_\epsilon^2\right) - \textstyle\sum_{t=2}^{n} \tfrac{(y_t - (\phi_0 + \phi_1 y_{t-1}))^2}{2\sigma_\epsilon^2} \end{aligned}$$

Stima di MV condizionata per un MA(1)

- Per una MA(1) si ha:

$$Y_t = \mu - \theta_1 \epsilon_{t-1} + \epsilon_t$$

  on $|\theta| < 1$ e $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$
- In tal caso i parametri da stimare sono μ, θ_1 e σ_ϵ^2
- Si ha inoltre che la distribuzione di Y_t condizionatamente all'informazione fino all'istante $(t-1)$, I_{t-1}, è

$$Y_t \mid I_{t-1} \sim \mathcal{N}\left(\mu - \theta_1 \epsilon_{t-1}; \sigma_\epsilon^2\right)$$

Stima di MV condizionata per un MA(1)

- Assumiamo $\epsilon_0 = E(\epsilon_0) = 0$
- La log-verosimiglianza per una MA(1) si può scrivere come:

$$\log L(\delta \mid y_1, ..., y_n) = \sum_{t=1}^{n} \log f(y_t \mid I_{t-1}, \delta)$$

e cioè:

$$\log L(\delta) \propto -\frac{n}{2} \log \sigma_\epsilon^2 + \frac{1}{2\sigma_\epsilon^2} \sum_{t=1}^{n} (y_t - (\mu - \theta_1 \epsilon_{t-1}))^2$$

- A differenza delle y_t, le ϵ_t non sono osservate.

Stima di MV condizionata per un MA(1)

- Se si assume $\epsilon_0 = 0$ è possibile ricavare le ϵ_t in maniera ricorsiva
- Dalla relazione $y_t = \epsilon_t - \theta_1 \epsilon_{t-1}$ si ricava

$$\begin{aligned}
\epsilon_1 &= y_1 - (-\theta_1 \epsilon_0) = y_1 + \theta_1 \epsilon_0 \\
\epsilon_2 &= y_2 - (-\theta_1 \epsilon_1) = y_2 + \theta_1 y_1 + \theta_1^2 \epsilon_0 \\
\epsilon_3 &= y_3 - (-\theta_1 \epsilon_2) = y_3 + \theta_1 y_2 + \theta_1^2 y_1 + \theta_1^3 \epsilon_0 \\
&..... \\
\epsilon_t &= y_t + \theta_1 y_{t-1} + \theta_1^2 y_{t-2} + ... + \theta_1^{t-1} y_1 + \theta_1^t \epsilon_0
\end{aligned}$$

- Ogni ϵ_t dipende quindi da tutti i valori passati delle y_t

Controllo diagnostico

- Dopo la stima dei parametri dobbiamo controllare l'adeguatezza del modello controllando se le assunzioni fatte sono soddisfatte
- Significatività dei parametri (statistiche t)
- Stazionarietà e invertibilità (radici delle equazioni caratteristiche)
- Parsimonia (criteri AIC, BIC, ...)
- Analisi dei residui

Analisi dei residui

- L'assunzione di base è che i residui, ϵ_t, sono WN (cioè a media nulla, varianza costante e incorrelati)
- Per ogni modello stimato, possiamo calcolare i residui stimati, $\hat{\epsilon}_t = y_t - \hat{y}_t$ che sono stime dei veri ϵ_t non osservabili
- Il controllo dell'adeguatezza del modello ai dati è, pertanto, effettuato attraverso un'attenta analisi della serie dei residui $\hat{\epsilon}_t$
- Per controllare se la varianza è costante possiamo esaminare il grafico dei residui

Analisi dei residui

- Un controllo fondamentale è quello relativo all'incorrelazione dei residui, infatti, se i residui risultano incorrelati vuol dire che abbiamo spiegato adeguatamente la struttura di dipendenza (lineare) presente nei dati.
- Funzioni di autocorrelazione empiriche (ACF e PACF) dei residui
- Test di Ljung-Box:

$$H_0 : \rho_1 = \rho_2 = \ldots = \rho_k = 0$$

dove la statistica test e la sua distribuzione (approssimata) sotto H_0 sono:

$$Q = n(n+2)\sum_{i=1}^{k} \frac{\hat{\rho}_i^2}{n-i} \sim \chi^2_{k-p-q}$$

Analisi dei residui

- Per verificare la casualità dei residui si può rappresentare il grafico dei residui al tempo t contro i residui al tempo $t-1$ e utilizzare altri test di casualità.
- Se si deve verificare anche l'ipotesi di normalità dei residui si può:
  - costruire l'istogramma dei residui standardizzati $\hat{\epsilon}_t/\hat{\sigma}_\epsilon$ e confrontarlo con quello della distribuzione Normale,
  - rappresentare il Q-Q plot rispetto alla distribuzione Normale,
  - utilizzare opportuni test di Normalità, ad esempio il testi di Jarque e Bera.

Test di Jarque e Bera

- Sia S il coefficiente di asimmetria e K il coefficiente di curtosi.
- Sotto ipotesi di normalità e asintoticamente si ha che:

$$\tilde{S} = \sqrt{\frac{n}{6}}\hat{S} \sim N(0,1)$$
$$\tilde{K} = \sqrt{\frac{n}{24}}(\hat{K}-3) \sim N(0,1)$$

- Statistica test di Jarque e Bera

$$JB = \tilde{S}^2 + \tilde{K}^2$$

- Sotto ipotesi di normalità e asintoticamente si ha che:

$$JB \sim \chi^2_2$$