Serie Storiche

La stima delle componenti di una serie storica

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

Stima delle componenti di una serie storica

$$\begin{aligned} Y_t &= f(t) + \varepsilon_t \\ &= f(T_t, C_t, S_t) + \varepsilon_t \end{aligned} \tag{1}$$

con $\varepsilon_t \sim WN(0, \sigma^2_\varepsilon)$

Nel corso non verrà trattata la stima di C_t ma verrà modellato il cosiddetto Trend-Ciclo.

Il modello (1) può essere stimato con:

- Metodi parametrici
  $\Longrightarrow$ $f(t)$ una funzione deterministica del tempo
- Metodi non parametrici
  $\Longrightarrow$ medie mobili

Approccio parametrico

A. **Solo trend** $f(t) = T_t \quad \Rightarrow Y_t = T_t + \varepsilon_t$
   - A.1 Trend lineare o linearizzabile $\Rightarrow$ regressione polinomiale
   - A.2 Trend non lineare e non linearizzabile (curve di crescita)

B. **Solo stagionalità** $f(t) = S_t \quad \Rightarrow Y_t = S_t + \varepsilon_t$
   - B.1 Regressione con variabili *dummy*
   - B.2 Regressione con funzioni trigonometriche

C. **Trend e stagionalità** $f(t) = T_t + S_t \quad \Rightarrow Y_t = T_t + S_t + \varepsilon_t$
   - C.1 Regressione (stima simultanea)

Trend lineare (nei parametri)

- Ipotizziamo di essere interessati alla sola stima del trend

$$Y_t = T_t + \varepsilon_t \quad \varepsilon_t \sim WN(0, \sigma^2_\varepsilon)$$

- Si può pensare di modellare il trend con una funzione polinomiale del tempo, ossia t

$$f(t) = T_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \ldots + \alpha_q t^q$$

- Modello

$$Y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \ldots + \alpha_q t^q + \varepsilon_t \quad t = 1, 2, \ldots, n$$

Modello in forma matriciale

$$Y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \ldots + \alpha_q t^q + \varepsilon_t, \quad t = 1, 2, \ldots, n$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & 1 & \ldots & 1^q \\ 1 & 2 & \ldots & 2^q \\ \vdots & \vdots & \ldots & \vdots \\ \vdots & \vdots & \ldots & \vdots \\ 1 & n & \ldots & n^q \end{bmatrix} \cdot \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_q \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\boldsymbol{y} = \boldsymbol{P}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$$

che è un modello di regressione multipla, pertanto, risolvendo:

$$\sum_{t=1}^{n} \left[y_t - (\alpha_0 + \alpha_1 t + \alpha_2 t^2 + \ldots + \alpha_q t^q) \right]^2 = \min_{\alpha}$$

si ottiene:

$$\hat{\alpha} = (\boldsymbol{P}'\boldsymbol{P})^{-1}\boldsymbol{P}'\boldsymbol{y}$$

Ordine q del polinomio

- L'ordine q del polinomio dipende dal comportamento di fondo della serie:

$$
\begin{aligned}
Y_t &= \alpha_0 + \varepsilon_t && \text{Trend costante} \\
Y_t &= \alpha_0 + \alpha_1 \cdot t + \varepsilon_t && \text{Trend lineare} \\
Y_t &= \alpha_0 + \alpha_1 \cdot t + \alpha_2 \cdot t^2 + \varepsilon_t && \text{Trend parabolico}
\end{aligned}
$$

- q non deve essere troppo elevato
- cautela per quanto attiene alle previsioni

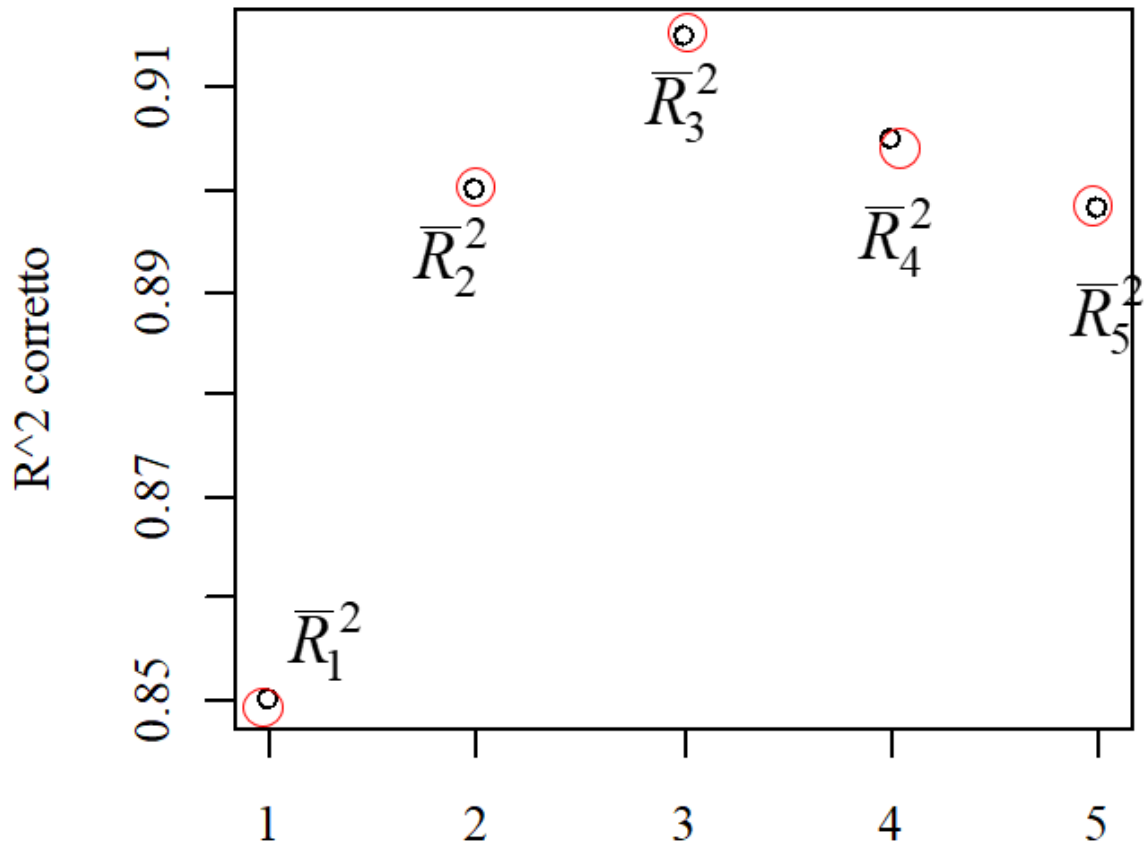
Criteri per la scelta di q

- Confronto tra $\bar{R}^2$
- Significatività dei parametri
- Criterio delle differenze successive (si veda il libro)

Confronto tra $\bar{R}^2$ corretto

- L'aggiunta di una nuova variabile esplicativa non provoca un aumento della devianza residua $\Rightarrow R_r^2 \leq R_{r+1}^2$
- La stessa relazione non vale per: $\bar{R}_r^2 = 1 - \left[\frac{n-1}{n-r-1}(1 - R_r^2)\right]$
- Confronto tra $\bar{R}_r^2$ e $\bar{R}_{r+1}^2$ per la scelta di q:
  - ci si arresta al grado r (quindi $q = r$) se $\bar{R}_r^2 \geq \bar{R}_{r+1}^2$
  - si prosegue nella ricerca se $\bar{R}_r^2 < \bar{R}_{r+1}^2$
- Significatività dei coefficienti associati alle potenze di t

Criterio dell'R^2 corretto



Il tasso di disoccupazione in Italia 1992.4 - 2001.2

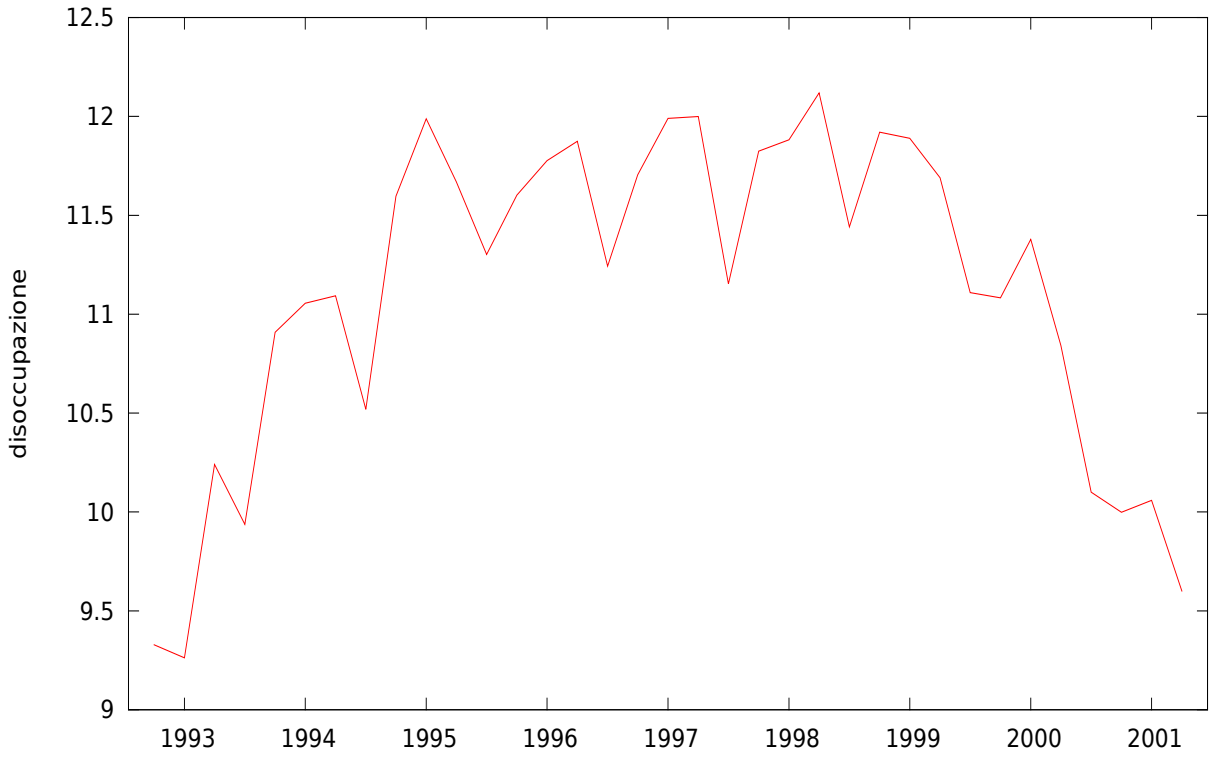


Figure: Esempio a pag. 44 del libro (file *figura2-1.txt*)

Scelta dell'ordine del polinomio

	$\bar{R}_r^2$	Significativ ità dei parametri

α_0	α_1	α_2	α_3
$r = 0$	$\bar{R}_0^2 = 0$	0.000	–	–	–
$r = 1$	$\bar{R}_1^2 = -0.019$	0.000	0.546	–	–
$r = 2$	$\bar{R}_2^2 = 0.823$	0.000	0.000	0.000	–
$r = 3$	$\bar{R}_3^2 = 0.817$	0.000	0.000	0.066	0.925

$$q = 2$$

Figure: In tabella sono riportati i $p-$ value

Si noti che se $\bar{R} \simeq 0$ allora $(1 - \bar{R}) \simeq 0$ e, dato che $(n-1)/(n-r-1) > 1$, $\bar{R}$ può diventare negativo

Il tasso di disoccupazione in Italia: trend quadratico

Modello 3: OLS, usando le osservazioni 1992:4-2001:2 (T = 35)
Variabile dipendente: disoccupazione

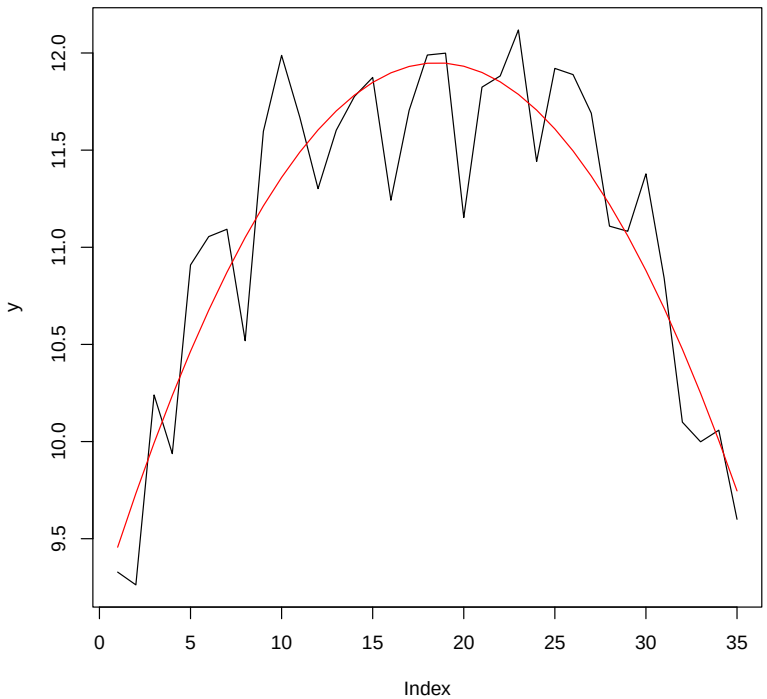
	coefficiente	errore std.	rapporto t	p-value	
const	9.16371	0.187244	48.94	1.17e-31	***
time	0.300766	0.0239828	12.54	6.80e-14	***
time2	-0.00811795	0.000646205	-12.56	6.49e-14	***

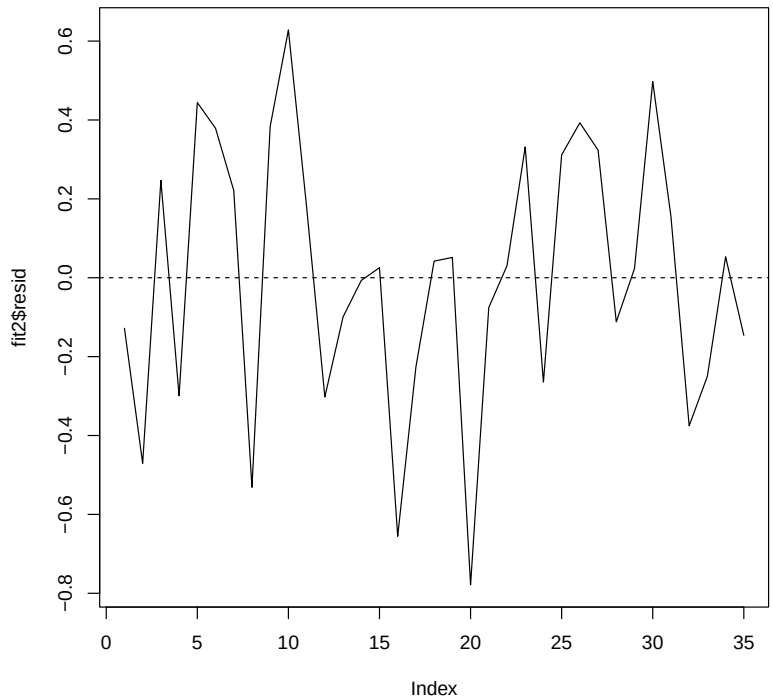
Media var. dipendente	11.11925	SQM var. dipendente	0.827700
Somma quadr. residui	3.883140	E.S. della regressione	0.348351
R-quadro	0.833291	R-quadro corretto	0.822872
F(2, 32)	79.97590	P-value(F)	3.56e-13
Log-verosimiglianza	-11.18553	Criterio di Akaike	28.37106
Criterio di Schwarz	33.03711	Hannan-Quinn	29.98178
rho	0.075791	Durbin-Watson	1.839525

Note: SQM = scarto quadratico medio; E.S. = errore standard

Il tasso di disoccupazione in Italia: l'adattamento ai dati

$$Y_t = 9.164 + 0.301 \cdot t - 0.008 \cdot t^2 + \varepsilon_t, \qquad \varepsilon_t \sim N(0, 0.348^2)$$





Esercizio

Per la serie del tasso di disoccupazione italiano, y_t,:

1. stimare il trend;
2. determinare quanto vale il trend nei trimestri 1993:3 e1994:2;
3. determinare quanto vale la componente residua negli stessi trimestri;
4. fare una previsione del trend per il trimestre 2001:3;

```
y = 9.328161 9.262356 10.23992 9.937655 10.90888 11.055 11.09276
10.51899 11.59662  11.98787 11.66901 11.30161 11.60235 11.7764
11.8741 11.24226  11.70456 11.98945 11.99912 11.15315 11.82424
11.88189 12.11866  11.44164 11.92071 11.88877011.6900 11.10875
11.08220 11.37809 10.84132 10.100  9.998741 10.05856  9.600
```

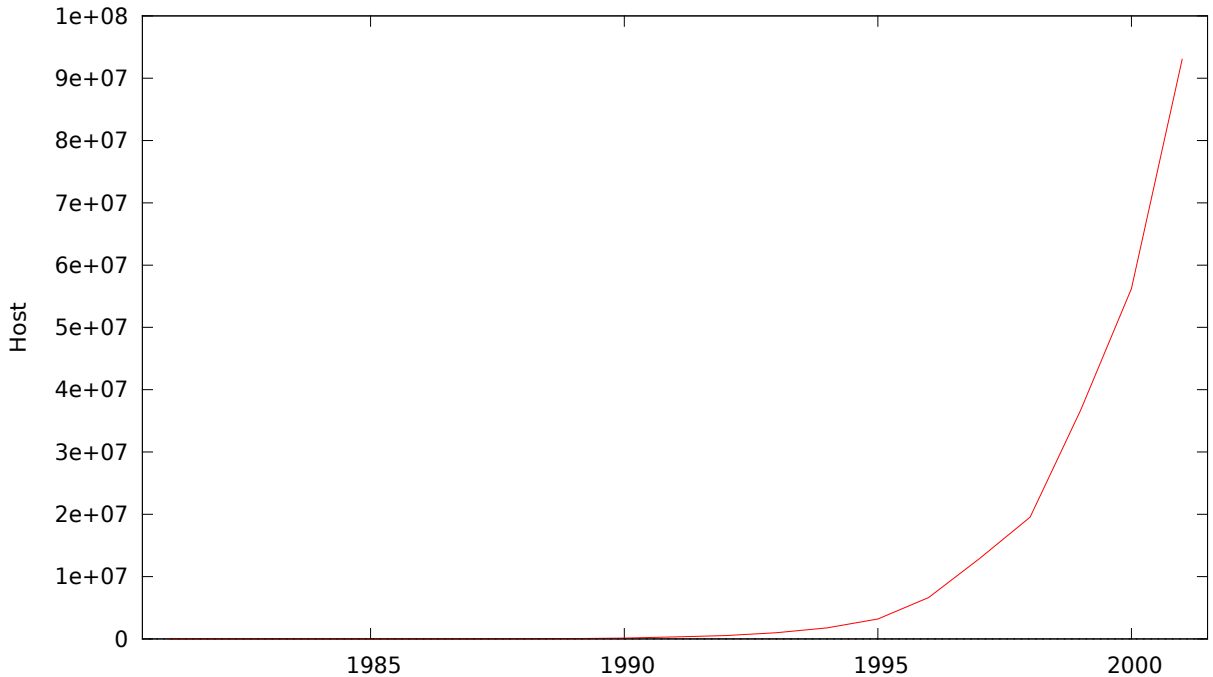
Soluzione

1. $\hat{T}_t = 9.164 + 0.301 \cdot t - 0.008 \cdot t^2$
2. Nella sequenza delle osservazioni i trimestri 1993:3 e 1994:2 occupano, rispettivamente, le posizioni 4 e 7. Dunque:
   - $\hat{T}_{1993:3} = \hat{T}_4 = 9.164 + 0.301 \cdot 4 - 0.008 \cdot 4^2 = 10.24$
   - $\hat{T}_{1994:2} = \hat{T}_7 = 9.164 + 0.301 \cdot 7 - 0.008 \cdot 7^2 = 10.879$
3. componente residua:
   - $\hat{\varepsilon}_{1993:3} = y_{1993:3} - \hat{T}_{1993:3} = 9.937655 - 10.24 = -0.302$
   - $\hat{\varepsilon}_{1994:2} = y_{1994:2} - \hat{T}_{1994:2} = 11.09276 - 10.879 = 0.21376$
4. Il trimestre 2001:3 corrisponde al tempo $t = 36$ e dunque:

$$\hat{T}_{2001:3} = \hat{T}_{36} = 9.164 + 0.301 \cdot 36 - 0.008 \cdot 36^2 = 9.632$$

Trend non lineari

- Per alcuni tipi di fenomeni un trend polinomiale può non essere adatto.
- Esempio: numero di server Internet nel mondo (1981 - 2001) (file *figura2-4.txt*)



Trend esponenziale

- Trend esponenziale

$$f(t) = \alpha_0 e^{\alpha_1 \cdot t}, \quad \alpha_0 > 0, \ \alpha_1$$

- Tasso di crescita direttamente proporzionale al livello della funzione:

$$\frac{df(t)}{dt} = \alpha_0 \cdot \alpha_1 e^{\alpha_1 \cdot t} = \alpha_1 \cdot f(t)$$

- Tasso di crescita relativo costante:

$$\frac{df(t)}{dt} / f(t) = \alpha_1$$

Trend esponenziale - modello moltiplicativo

- Assumiamo che

$$Y_t = f(t) \cdot \varepsilon_t = \alpha_0 e^{\alpha_1 \cdot t} \cdot \varepsilon_t$$

- Modello linearizzato

$$\begin{aligned} \log Y_t &= \log f(t) + \log \varepsilon_t \\ &= \log \alpha_0 + \alpha_1 \cdot t + \log \varepsilon_t \\ &= \alpha_0^* + \alpha_1 \cdot t + \varepsilon_t^* \end{aligned}$$

- Se Y_t segue un trend esponenziale, $\log Y_t$ segue un trend lineare.
- Possiamo stimare i parametri del modello linearizzato
- Si osservi che, se $\varepsilon_t^* \sim N(0, \sigma^2)$ allora

$$\varepsilon_t \sim LogN(e^{\sigma^2/2}, e^{\sigma^2}(e^{\sigma^2} - 1))$$

Trend esponenziale - modello moltiplicativo

- In questo caso però $E(Y_t) \neq f(t)$ e dunque la stima del trend è distorta, infatti

$$\begin{aligned} E[Y_t] &= E[f(t) \cdot \varepsilon_t] \\ &= E[\alpha_0 e^{\alpha_1 \cdot t} \cdot \varepsilon_t] \\ &= \alpha_0 e^{\alpha_1 \cdot t} \cdot E[\varepsilon_t] \\ &= \alpha_0 e^{\alpha_1 \cdot t} \cdot e^{\sigma^2/2} \neq \alpha_0 e^{\alpha_1 \cdot t} = f(t) \end{aligned}$$

  e dunque si ottiene una stima distorta del trend

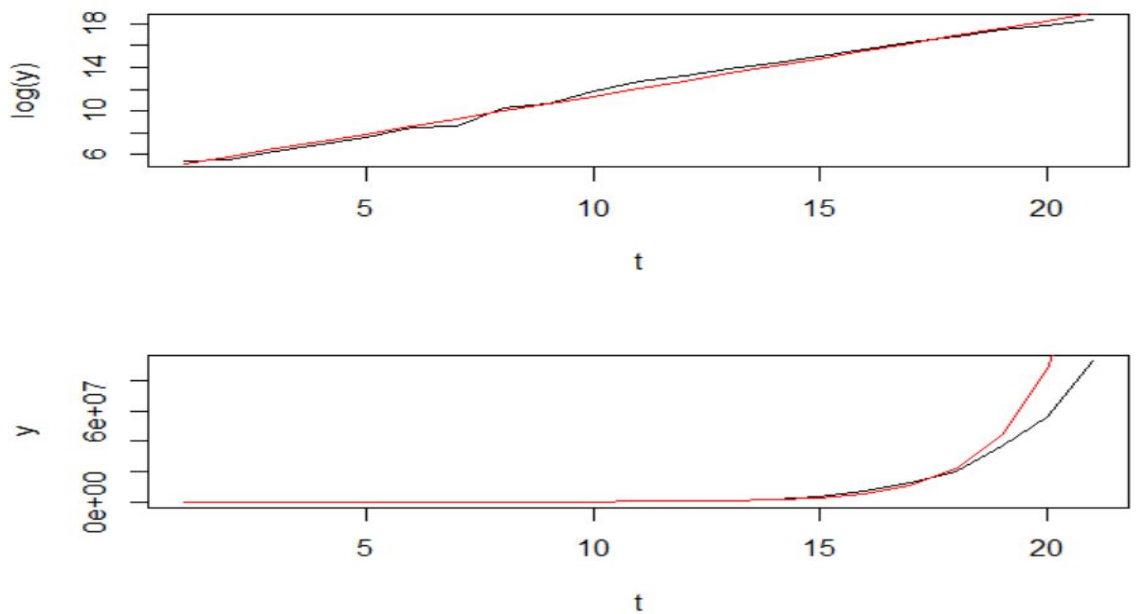
- Inoltre, mentre $\hat{\alpha}_1$ è stima OLS di α_1, $\hat{\alpha}_0$ è la stima OLS di $\alpha_0^* = \log \alpha_0$ per cui

$$\hat{\alpha}_0 = \exp(\hat{\alpha_0^*}) = \exp(\widehat{\log \alpha_0})$$

  ma questa non è più stima OLS di α_0.

Il numero di server internet nel mondo, trend lineare: $\log f(t) = \alpha_0 + \alpha_1 \cdot t$

	coefficiente	errore std.	rapporto t	p-value
const	4.41078	0.170997	25.79	2.99e-16 ***
time	0.693618	0.0136182	50.93	8.78e-22 ***



Trend esponenziale - modello additivo

- Se assumiamo che

$$Y_t = f(t) + \varepsilon_t = \alpha_0 e^{\alpha_1 \cdot t} + \varepsilon_t$$

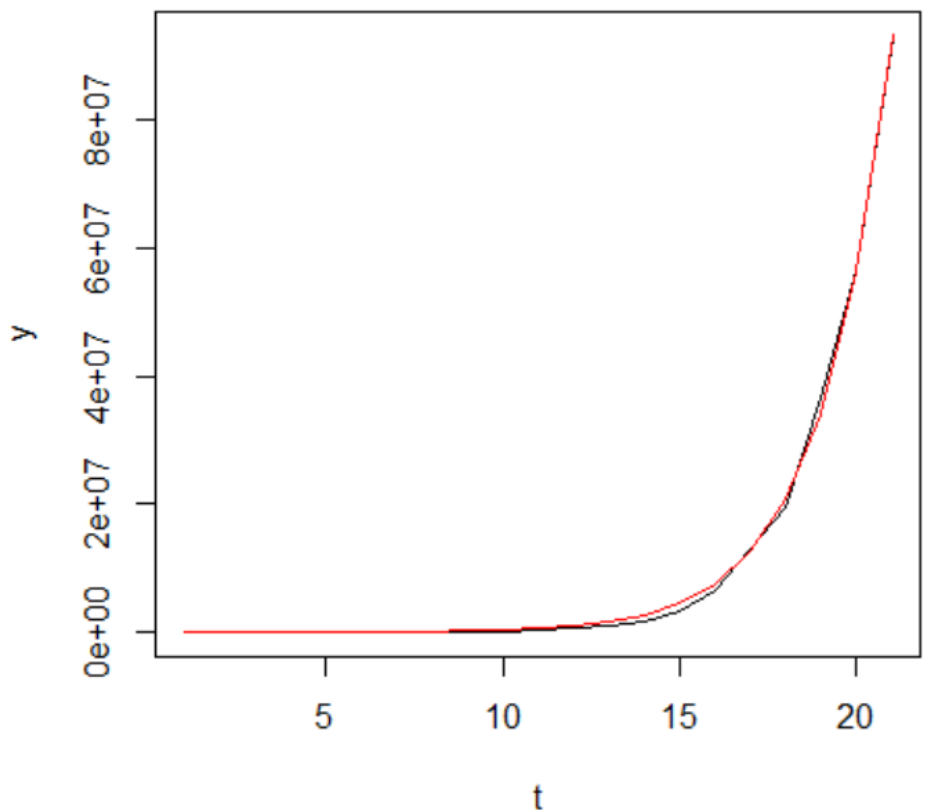
  otteniamo un modello non lineare e non linearizzabile
- Stima dei parametri mediante minimi quadrati non lineari, minimizzando

$$S(\alpha_0, \alpha_1) = \sum_{t=1}^{T} \left(Y_t - \alpha_0 e^{\alpha_1 \cdot t} \right)^2$$

- In questo caso non esiste una soluzione analitica e la minimizzazione richiede l'utilizzo di metodi numerici.

Il numero di server internet nel mondo, trend esponenziale: $f(t) = \alpha_0 \cdot e^{\alpha_1 \cdot t}$

```
Formula: y ~ a0 * exp(a1 * tempo)
Parameters:
   Estimate Std. Error t value Pr(>|t|)
a0 2.423e+03  3.633e+02    6.67 2.23e-06
a1 5.028e-01  7.333e-03   68.58  < 2e-16
```

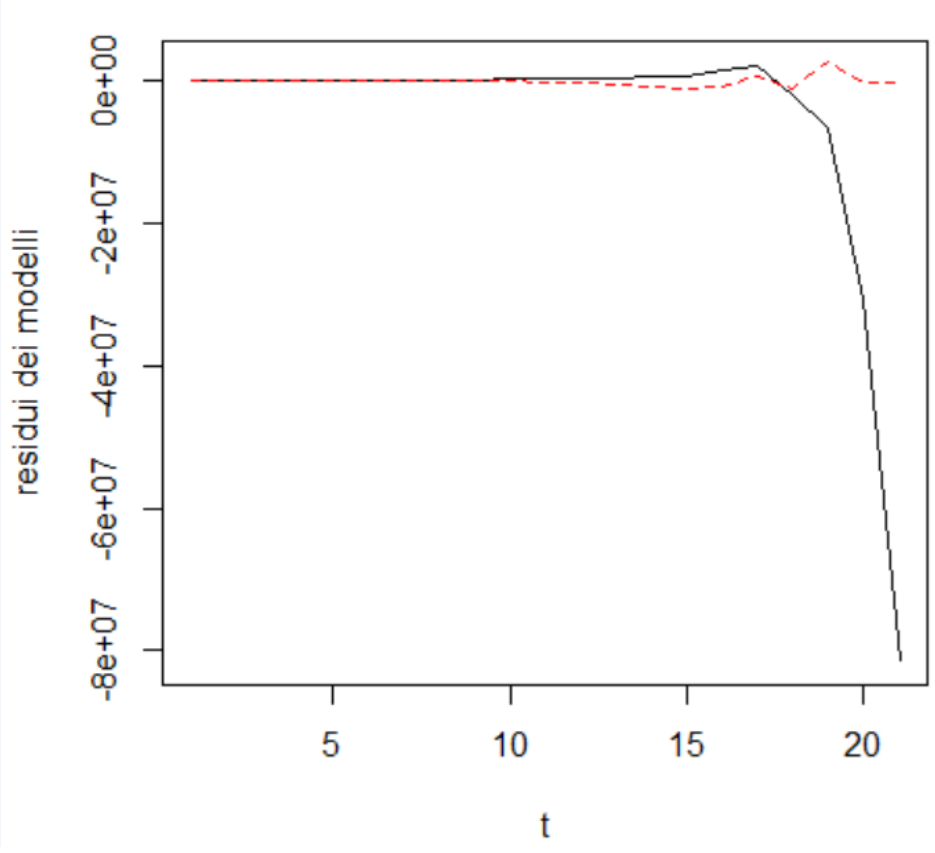


Il numero di server internet nel mondo, trend esponenziale: $f(t) = \alpha_0 \cdot e^{\alpha_1 \cdot t}$

- Comandi R per la stima non lineare

```
tempo=seq(1,length(y),1)
nlmod=nls(y ~ a0 *exp(a1*tempo), start = list(a0
= 1000, a1 = 0.8))
summary(nlmod)
```

Residui dei due modelli



Residui dei modelli. Linea nera continua: modello moltiplicativo; linea rossa tratteggiata: modello additivo.

Esercizio

1. Fornire una stima del trend per gli anni 1994 e 1995.
2. Stimare il valore dei residui negli stessi anni.
3. Fornire una previsione del trend per il 2005.

- Trend lineare stimato:

$$\log \hat{f}(t) = \log \hat{\alpha}_0 + \hat{\alpha}_1 \cdot t = \hat{\alpha}_0^* + \hat{\alpha}_1 \cdot t = 4.411 + 0.694 \cdot t$$

quindi:

$$\hat{f}(t) = e^{\hat{\alpha}_0^*} \cdot e^{\hat{\alpha}_1 \cdot t} = 82.352 \cdot e^{0.694 \cdot t}$$

Soluzione col modello linearizzato

1. 1994 e 1995 corrispondono a $t = 14$ e $t = 15$, pertanto:

$$\hat{T}_{1994} = \hat{T}_{14} = 82.352 \cdot e^{0.694 \cdot 14} = 1365461$$

$$\hat{T}_{1995} = \hat{T}_{15} = 82.352 \cdot e^{0.694 \cdot 15} = 2733252$$

2. residui:

$$\hat{\varepsilon}_{1994} = y_{1994} - \hat{T}_{1994} = 1.766.000 - 1.365.461 = 400.539$$

$$\hat{\varepsilon}_{1995} = y_{1995} - \hat{T}_{1995} = 3.212.000 - 2.733.252 = 478.748$$

3. il 2005 corrisponde a t=25 e dunque:

$$\hat{T}_{2005} = \hat{T}_{25} = 82.352 \cdot e^{0.694 \cdot 25} = 282.821.470$$

Soluzione col modello non lineare

1. Trend stimato

$$f(\hat{t}) = \hat{\alpha}_0 e^{\hat{\alpha}_1 t} = 2423.37 e^{0.502845} t$$

1. pertanto:

$$\hat{T}_{1994} = \hat{T}_{14} = 2423.37 \cdot e^{0.502845 \cdot 14} = 2765528$$

$$\hat{T}_{1995} = \hat{T}_{15} = 2423.37 \cdot e^{0.502845 \cdot 15} = 4572574$$

2. residui:

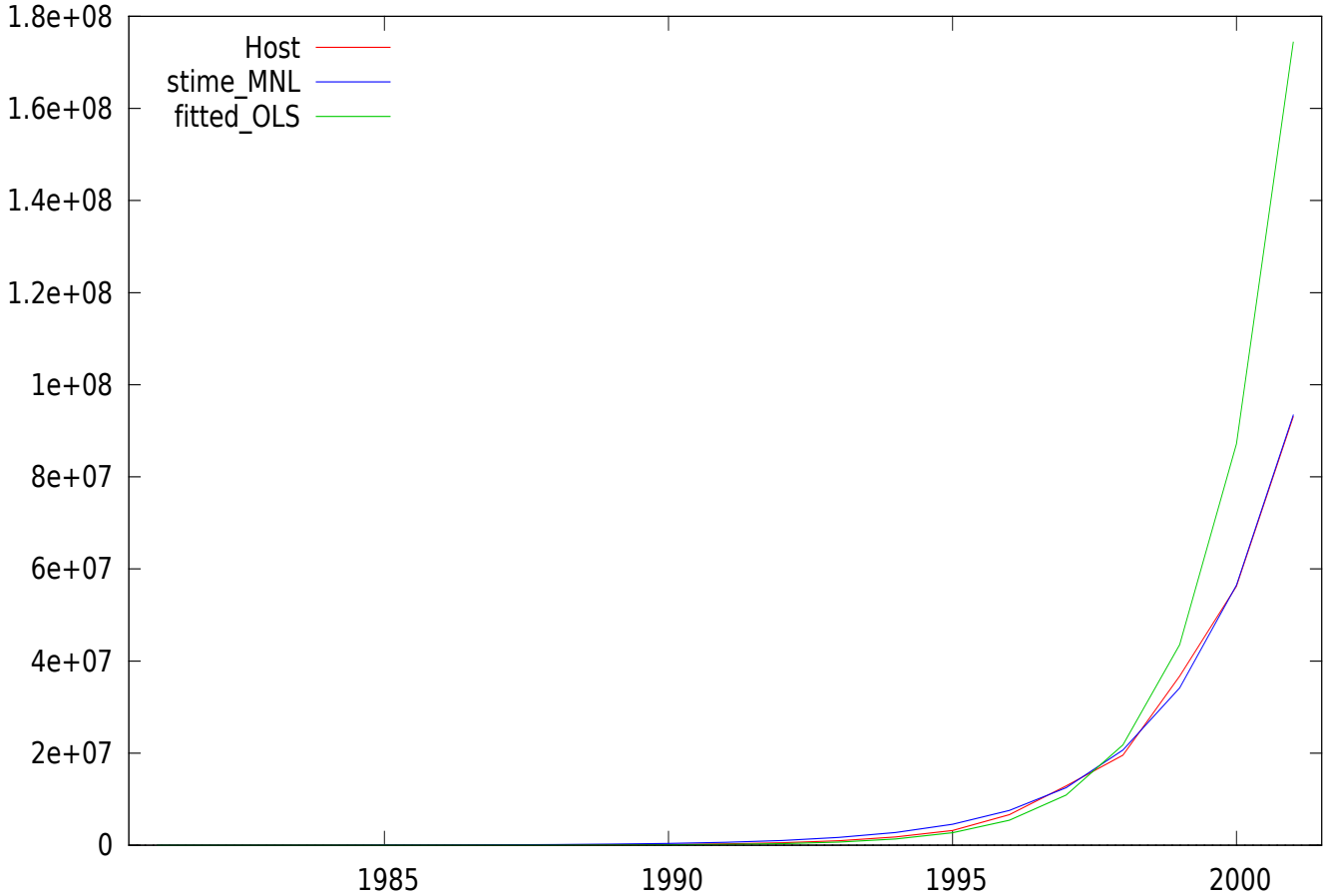
$$\hat{\varepsilon}_{1994} = y_{1994} - \hat{T}_{1994} = 1766000 - 2765528 = -999528$$

$$\hat{\varepsilon}_{1995} = y_{1995} - \hat{T}_{1995} = 3212000 - 4572574 = -1360574$$

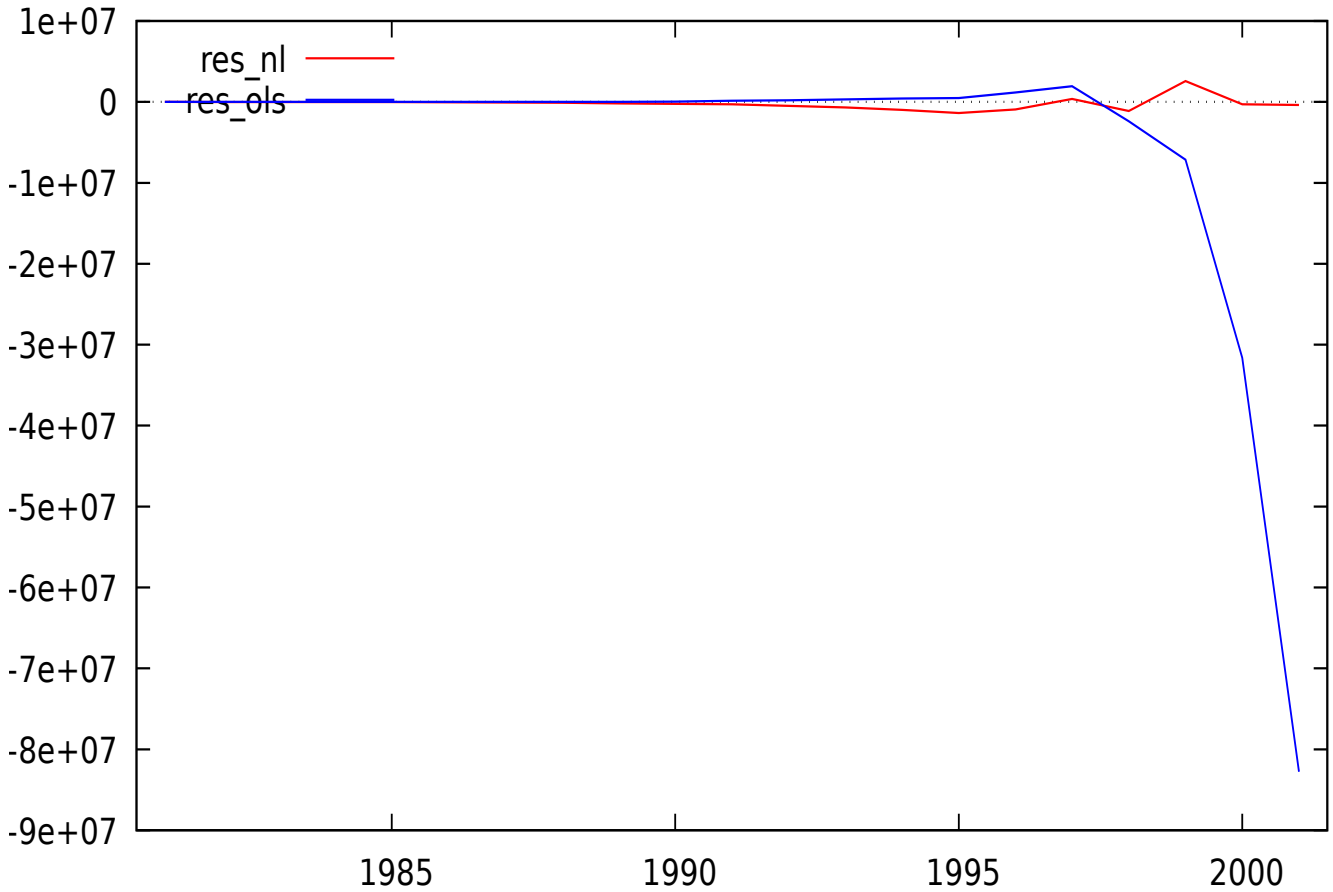
3. il 2005 corrisponde a t=25 e dunque:

$$\hat{T}_{2005} = \hat{T}_{25} = 2423.37 \cdot e^{0.502845 \cdot 25} = 698216240$$

Confronto fra valori stimati col modello lineare e con quello non lineare



Confronto fra residui del modello lineare e non lineare



La componente stagionale

- La componente stagionale, S_t, può essere rappresentata da una funzione periodica, $g(t)$:

$$g(t) = g(t+s) = g(t+2s) = \ldots$$

- le oscillazioni (tutte della stessa forma) si ripetono con periodo s uguale e ampiezza costante.

Esempio

- per serie trimestrali $s = 4$
- per serie mensili $s = 12$
- ...

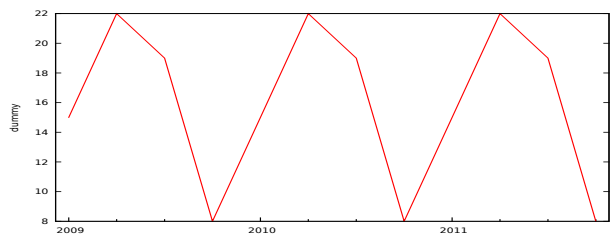
Rappresentazione di una funzione periodica mediante variabili *dummy*

Supponiamo $Y_t = S_t = g(t), \quad t = 1, \ldots, 12 \quad s = 4$. Allora:

$$g(t) = \sum_{j=1}^{4} \gamma_j d_{jt} = \gamma_1 d_{1t} + \gamma_2 d_{2t} + \gamma_3 d_{3t} + \gamma_4 d_{4t}, \quad t = 1, 2, \ldots, 12$$

$$d_{jt} = \begin{cases} 1 & \text{nel periodo } j\text{–esimo dell'anno a cui appartiene } t \\ 0 & \text{altrimenti} \end{cases}$$

d_{jt} vengono chiamate variabili *dummy*.



Esempio con $\gamma_1 = 15$, $\gamma_2 = 22$, $\gamma_3 = 19$ e $\gamma_4 = 9$.

Rappresentazione matriciale

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ \vdots \\ y_{12} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \\ \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_4 \end{bmatrix} \tag{2}$$

che in forma matriciale diventa:

$$\boldsymbol{y} = \boldsymbol{D}\gamma \tag{3}$$

Coefficienti di stagionalità

Più in generale, assumendo $Y_t = S_t + \varepsilon_t$, si ha il modello di regressione multipla:

$$\boldsymbol{y} = \boldsymbol{D}\gamma + \varepsilon$$

o, equivalentemente:

$$y_t = \gamma_1 d_{1t} + \gamma_2 d_{2t} + \gamma_3 d_{3t} + \gamma_4 d_{4t} + \varepsilon_t, \quad t = 1, 2, \ldots, n.$$

I coefficienti stagionali da stimare sono γ_1, γ_2, γ_3 e γ_4.
Indichiamo con $\hat{\gamma}^*$ la loro stima OLS:

$$\hat{\gamma}^* = (\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{y}$$

I coefficienti $\hat{\gamma}_j{}^* = (\gamma_1^*, \gamma_2^*, \gamma_3^*, \gamma_4^*)$ sono detti **coefficienti grezzi di stagionalità** e non sono ancora i coefficienti stagionali finali.

Modello scritto in termini scalari

$$
\begin{aligned}
y_1 &= \gamma_1 d_{11} + \gamma_2 d_{21} + \gamma_3 d_{31} + \gamma_4 d_{41} + \varepsilon_1 \\
y_2 &= \gamma_1 d_{12} + \gamma_2 d_{22} + \gamma_3 d_{32} + \gamma_4 d_{42} + \varepsilon_2 \\
y_3 &= \gamma_1 d_{13} + \gamma_2 d_{23} + \gamma_3 d_{33} + \gamma_4 d_{43} + \varepsilon_3 \\
y_4 &= \gamma_1 d_{14} + \gamma_2 d_{24} + \gamma_3 d_{34} + \gamma_4 d_{44} + \varepsilon_4 \\
\ldots &= \ldots \\
\ldots &= \ldots \\
\ldots &= \ldots \\
y_{12} &= \gamma_1 d_{1,12} + \gamma_2 d_{2,12} + \gamma_3 d_{3,12} + \gamma_4 d_{4,12} + \varepsilon_{12}
\end{aligned}
$$

Coefficienti di stagionalità

- I coefficienti grezzi di stagionalità $\hat{\gamma}_j^*$ rappresentano solo una stima preliminare della componente stagionale:

$$\boldsymbol{D}\hat{\gamma}^* \Rightarrow \hat{g}(t) = \sum_{j=1}^{4} \hat{\gamma}_j^* d_{jt} \quad t = 1, 2, \ldots, n$$

- Per poter parlare di componente stagionale bisogna che essa si esaurisca nell'arco del periodo considerato (anno, settimana, etc).
  Per un modello additivo ciò equivale a chiedere che:

$$\sum_{j=1}^{s} \hat{\gamma}_j^* = 0$$

Perché vogliamo che i coefficienti stagionali abbiano somma nulla?

- Consideriamo un modello del tipo:

$$y_t = \gamma_1 d_{1t} + \gamma_2 d_{2t} + \gamma_3 d_{3t} + \gamma_4 d_{4t} + \varepsilon_t \tag{4}$$

- Se non c'è alcuna componente stagionale, il livello di ogni stagione (trimestre) è sempre lo stesso; se lo indichiamo con $\bar{\gamma}$ si può scrivere:

$$y_t = \bar{\gamma} d_{1t} + \bar{\gamma} d_{2t} + \bar{\gamma} d_{3t} + \bar{\gamma} d_{4t} + \varepsilon_t = \bar{\gamma} + \varepsilon_t \tag{5}$$

- Se c'è una componente stagionale il modello (4) può essere scritto anche in modo da evidenziare lo scostamento di ogni stagione dalla media sull'intero periodo:

$$y_t = \bar{\gamma} + (\gamma_1 - \bar{\gamma}) d_{1t} + (\gamma_2 - \bar{\gamma}) d_{2t} + (\gamma_3 - \bar{\gamma}) d_{3t} + (\gamma_4 - \bar{\gamma}) d_{4t} + \varepsilon_t \tag{6}$$

$$y_t = \bar{\gamma} + (\gamma_1 - \bar{\gamma})d_{1t} + (\gamma_2 - \bar{\gamma})d_{2t} + (\gamma_3 - \bar{\gamma})d_{3t} + (\gamma_4 - \bar{\gamma})d_{4t} + \varepsilon_t$$

- Se per tutti i trimestri $\gamma_i = \bar{\gamma}$ si torna al modello (5)
- Al fine della destagionalizzazione quello che conta è la differenza $\gamma_i - \bar{\gamma}$ perché questa rappresenta la deviazione della specifica stagione dal livello medio dell'intero periodo
- Quando si aggregano i dati si ottiene una cosa del tipo:

$$y_t^{t+3} = \sum_{i=t}^{t+3} y_i = 4\bar{\gamma} + \sum_{i=1}^{4}(\gamma_i - \bar{\gamma}) + \sum_{i=t}^{t+3} \varepsilon_i$$

quindi affinché la componente stagionale si annulli deve essere $\sum_{i=1}^{4}(\gamma_i - \bar{\gamma}) = \sum_{i=1}^{4} \hat{\gamma}_i = 0$.

Coefficienti ideali di stagionalità, γ_j

- La condizione di somma nulla in generale non è soddisfatta, tuttavia è sempre vera per gli scarti dalla media.
- Sia dunque

$$\bar{\gamma} = \frac{1}{s} \sum_{j=1}^{s} \hat{\gamma}_j^*$$

la media degli s coefficienti stagionali grezzi.

- Chiamiamo **coefficienti ideali di stagionalità**

$$\hat{\gamma}_j = \hat{\gamma}_j^* - \bar{\gamma} \qquad j = 1, 2, \ldots, s$$

gli scarti dei coefficienti grezzi dalla loro media.

Serie destagionalizzata

- La destagionalizzazione deve avvenire in usando i coefficienti ideali di stagionalità.
- Stima della (*vera*) componente stagionale:

$$\boldsymbol{D}\hat{\gamma} \Rightarrow \hat{g}(t) = \sum_{j=1}^{4} \hat{\gamma}_j d_{jt} \quad t = 1, 2, \ldots, n$$

- Serie destagionalizzata

$$y_t^d = y_t - \hat{g}(t) = y_t - \sum_{j=1}^{s} \hat{\gamma}_j d_{jt} \qquad t = 1, 2, \ldots, n$$

o, in forma matriciale: $\boldsymbol{y^d} = \boldsymbol{y} - \boldsymbol{D}\hat{\gamma}$

La trappola delle dummy

- Se scriviamo il modello inserendo anche la costante

$$y_t = \gamma_0 + \gamma_1 d_{1t} + \gamma_2 d_{2t} + \gamma_3 d_{3t} + \gamma_4 d_{4t} + \varepsilon_t, \quad t = 1, 2, \ldots, n.$$

- la matrice di regressione diventerebbe singolare!

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ \vdots \\ y_{12} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \vdots \\ \varepsilon_{12} \end{bmatrix} \tag{7}$$

Le "soluzioni" alla trappola delle dummy

1. Non includere mai la costante;
2. Includere la costante ma eliminare una (qualsiasi) delle variabili dummy. In questo caso i coefficienti hanno una diversa interpretazione.

   Se, ad esempio, si elimina la dummy del primo trimestre si avrà il modello

$$y_t = \alpha + \gamma_2 d_{2t} + \gamma_3 d_{3t} + \gamma_4 d_{4t} + \varepsilon_t, \quad t = 1, 2, \ldots, n.$$

   Indichiamo le stime OLS dei parametri del modello precedente con la tilde:

$$\tilde{\alpha},\ \tilde{\gamma}_2,\ \tilde{\gamma}_3,\ \tilde{\gamma}_4$$

Modello con la costante: interpretazione dei parametri

$$\tilde{\alpha} = \hat{\gamma}_1^*$$

$$\tilde{\gamma}_i = \hat{\gamma}_i^* - \tilde{\alpha} = \hat{\gamma}_i^* - \hat{\gamma}_1^* \qquad i = 2, 3, 4$$

da cui si ricava

$$\hat{\gamma}_i^* = \tilde{\gamma}_i + \tilde{\alpha}$$

- Dunque la stima della costante corrisponde alla stima del coefficiente grezzo di stagionalità della dummy eliminata, mentre gli altri coefficienti rappresentano effetti differenziali rispetto al (in questo caso) coefficiente del primo trimestre.

Modello con la costante: destagionalizzazione

- Se i coefficienti individuati ($\hat{\gamma}_i^*$) non sommano a zero bisogna considerare i coefficienti ideali.
- Una volta ottenuti i coefficienti ideali, come descritto nelle slide precedenti, la destagionalizzazione della serie avviene al solito modo.

Il tasso di disoccupazione in Italia: stima della componente stagionale

Se per la serie della disoccupazione ipotizziamo un semplice modello con la sola componente stagionale:

$$Y_t = S_t + \varepsilon_t = \gamma_1 d_{1t} + \gamma_2 d_{2t} + \gamma_3 d_{3t} + \gamma_4 d_{4t} + \varepsilon_t$$

otteniamo le seguenti stime:

	coefficiente	errore std.	rapporto t	p-value
dq1	11.2532	0.283537	39.69	3.83e-28
dq2	11.2361	0.283537	39.63	4.01e-28
dq3	10.8505	0.300736	36.08	6.92e-27
dq4	11.1074	0.283537	39.17	5.69e-28

che portano al modello (stimato):

$$\hat{y}_t = 11.2532 d_{1t} + 11.2361 d_{2t} + 10.8505 d_{3t} + 11.1074 d_{4t}$$

Abbiamo pertanto che:

$$\sum_{j=1}^{4} \hat{\gamma}_j^* = 11.2532 + 11.2361 + 10.8505 + 11.1074 = 44.4472$$

essendo:

$$\bar{\gamma} = \frac{1}{4} \sum_{j=1}^{4} \hat{\gamma}_j^* = 11.1118$$

i coefficienti *veri* di stagionalità risultano essere:

$$\begin{aligned}
\hat{\gamma}_1 &= 11.2532 - 11.1118 = 0.1414 \\
\hat{\gamma}_2 &= 11.2361 - 11.1118 = 0.1243 \\
\hat{\gamma}_3 &= 10.8505 - 11.1118 = -0.2613 \\
\hat{\gamma}_4 &= 11.1074 - 11.1118 = -0.0044
\end{aligned}$$

Il modello stimato può dunque scriversi:

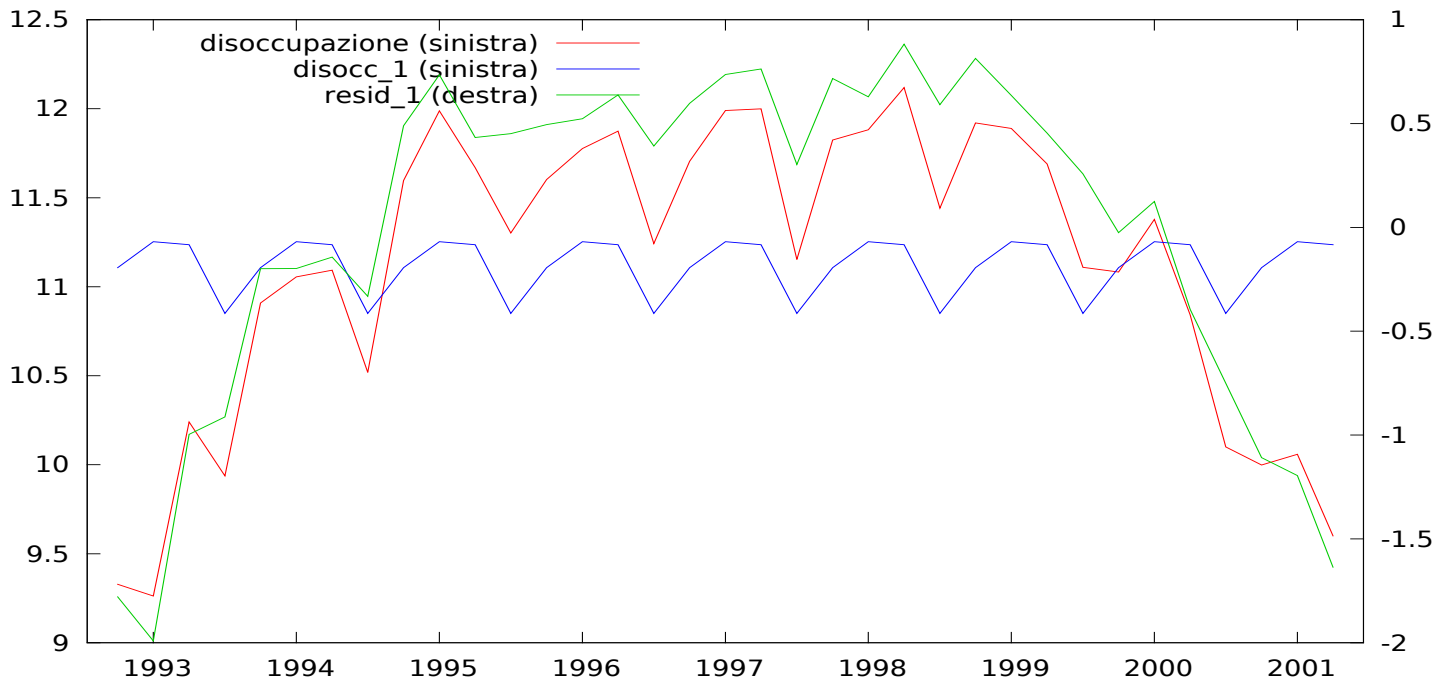
$$\hat{y}_t = 11.1118 + 0.1414 d_{1t} + 0.1243 d_{2t} - 0.2613 d_{3t} - 0.0044 d_{4t} \quad (8)$$

In questo caso i coefficienti delle dummy rappresentano le deviazioni di ogni trimestre dal livello medio.

Si noti che nel modello (8) sono presenti sia le quattro dummy sia la costante ma, in questo caso, la costante non è un parametro stimato.

Il tasso di disoccupazione in Italia: stima della stagionalità

Guardiamo il grafico relativo ai valori stimati e ai residui del modello contenente solo la componente stagionale:



Un vero disastro!!!

Stima simultanea di trend e stagionalità

- Se assumiamo più realisticamente che:

$$Y_t = T_t + S_t + \varepsilon_t \tag{9}$$

- il modello di regressione diventa:

$$\boldsymbol{y} = \boldsymbol{P}\boldsymbol{\alpha} + \boldsymbol{D}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{10}$$

$$\boldsymbol{y} = [\boldsymbol{P} \mid \boldsymbol{D}] \cdot \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\gamma} \end{bmatrix} + \boldsymbol{\varepsilon} = \boldsymbol{Z}\boldsymbol{\theta} + \boldsymbol{\varepsilon} \tag{11}$$

- da cui si ottiene lo stimatore

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y} \tag{12}$$

In forma matriciale

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & 1^2 & \dots & 1^q \\ 2 & 2^2 & \dots & 2^q \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ n & n^2 & \dots & n^q \end{bmatrix} \cdot \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_q \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$(\boldsymbol{y} = \boldsymbol{P}\boldsymbol{\alpha} + \boldsymbol{D}\boldsymbol{\gamma} + \varepsilon)$$

Proviamo a considerare un modello che tenga conto simultaneamente del trend e della componente stagionale di periodo $s = 4$.

Abbiamo già osservato che l'ordine del trend dovrebbe essere $q = 2$, e pertanto ipotizziamo il modello:

$$y_t = \alpha_1 t + \alpha_2 t^2 + \gamma_1 d_{1t} + \gamma_2 d_{2t} + \gamma_3 d_{3t} + \gamma_4 d_{4t} + \varepsilon_t \qquad t = 1, 2, \ldots, n$$

N.B. Si noti che il modello è senza costante per evitare la trappola delle dummy

Modello in forma matriciale

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & 1^2 \\ 2 & 2^2 \\ \vdots & \vdots \\ \vdots & \vdots \\ n & n^2 \end{bmatrix} \cdot \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$y_t = \alpha_1 t + \alpha_2 t^2 + \gamma_1 d_{1t} + \gamma_2 d_{2t} + \gamma_3 d_{3t} + \gamma_4 d_{4t} + \varepsilon_t \qquad t = 1, 2, \ldots, n$$

Se stimiamo il modello con gli OLS otteniamo:

	coefficiente	errore std.	rapporto t	p-value	
time	0.310134	0.0181319	17.10	1.09e-16	***
time2	-0.00838603	0.000488650	-17.16	9.94e-17	***
dq1	9.28232	0.158603	58.53	1.17e-31	***
dq2	9.26542	0.160518	57.72	1.74e-31	***
dq3	8.68959	0.169065	51.40	4.87e-30	***
dq4	9.15318	0.156127	58.63	1.11e-31	***

Media var. dipendente	11.11925	SQM var. dipendente	0.827700
Somma quadr. residui	1.988963	E.S. della regressione	0.261887
R-quadro	0.914611	R-quadro corretto	0.899889
F(5, 29)	62.12452	P-value(F)	1.33e-14
Log-verosimiglianza	0.522508	Criterio di Akaike	10.95498
Criterio di Schwarz	20.28707	Hannan-Quinn	14.17642
rho	0.437958	Durbin-Watson	1.112191

Note: SQM = scarto quadratico medio; E.S. = errore standard

Se stimiamo il modello con gli OLS otteniamo:

- I coefficienti stagionali stimati con OLS sono grezzi.
- Per calcolare i coefficienti di stagionalità ideali...
- Media dei coefficienti grezzi:

$$\bar{\gamma} = \frac{1}{4}\sum_{j=1}^{4} \hat{\gamma}_j^* = 9.097627$$

- Coefficienti di stagionalità veri

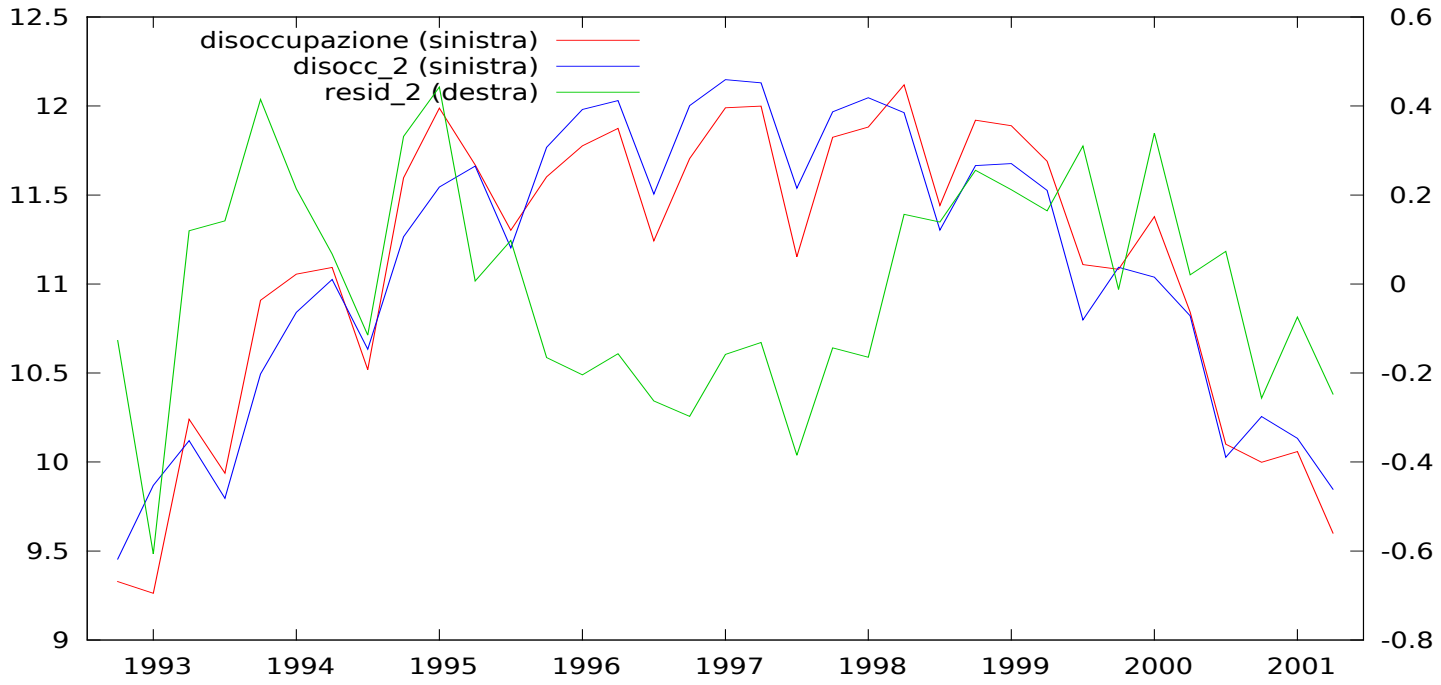
$$\begin{aligned}
\hat{\gamma}_1 &= 9.28232 - 9.097627 = 0.1846925 \\
\hat{\gamma}_2 &= 9.26542 - 9.097627 = 0.1677925 \\
\hat{\gamma}_3 &= 8.68959 - 9.097627 = -0.4080375 \\
\hat{\gamma}_4 &= 9.15318 - 9.097627 = 0.0555525
\end{aligned}$$

- Modello stimato

$$y_t = 9.0976 + 0.310134 \cdot t - 0.0083860 \cdot t^2 + 0.1846925 d_{1t}$$

$$+0.1677925 d_{2t} - 0.4080375 d_{3t} + 0.0555525 d_{4t}$$

Grafico valori stimati e residui



molto meglio!

Esercizio 2.3 pag. 9 dell'eserciziario

Alla serie storica trimestrale $y_t,\ t = 1, \ldots, 60$ è stato adattato il modello:

$$y_t = 13.5d_{1t} + 2.5d_{2t} - 5.5d_{3t} - 10.5d_{4t} + 0.5t + \varepsilon_t, \quad \varepsilon_t \sim WN(0, 3.2)$$

i. a quanto ammonta la variazione della serie nell'arco di un anno dovuta alla componente tendenziale? E quella dovuta alla componente stagionale?

ii. Si consideri la trasformazione $y_t^* = (1 - B^4)y_t = y_t - y_{t-4}$. Quanto vale mediamente la serie y_t^*?

iii. Si supponga che i valori $y_1 = 16$, $y_2 = 4$, $y_3 = -2$ e $y_4 = -9$ siano relativi ai quattro trimestri del primo anno. Si calcolino i corrispondenti valori destagionalizzati.

iv. Fornire delle previsioni per i valori assunti da y_t nei quattro trimestri successivi al periodo di osservazione.

Soluzione

i. a quanto ammonta la variazione della serie nell'arco di un anno dovuta alla componente tendenziale?

$$T_t - T_{t-4} = 0.5 \cdot t - 0.5 \cdot (t-4) = 2$$

E quella dovuta alla componente stagionale?

$$S_t - S_{t-4} = 0$$

ii. Si consideri la trasformazione $y_t^* = (1 - B^4)y_t = y_t - y_{t-4}$. Quanto vale mediamente la serie y_t^*?

$$\begin{aligned} y_t^* &= y_t - y_{t-4} = \\ &= 13.5d_{1t} + 2.5d_{2t} - 5.5d_{3t} - 10.5d_{4t} + 0.5t + \varepsilon_t - (13.5d_{1,t-4} + \\ &\quad +2.5d_{2,t-4} - 5.5d_{3,t-4} - 10.5d_{4,t-4} + 0.5(t-4) + \varepsilon_{t-4}) = \\ &= 2 + \varepsilon_t - \varepsilon_{t-4} \end{aligned}$$

pertanto: $E[y_t^*] = E[2 + \varepsilon_t - \varepsilon_{t-4}] = 2$.

Soluzione

iii. Si supponga che i valori $y_1 = 16$, $y_2 = 4$, $y_3 = -2$ e $y_4 = -9$ siano relativi ai quattro trimestri del primo anno. Si calcolino i corrispondenti valori destagionalizzati.
Osserviamo che:

$$\sum_{j=1}^{4} \hat{\gamma}_j^* = 13.5 + 2.5 - 5.5 - 10.5 = 0$$

e dunque i coefficienti non necessitano di alcuna trasformazione. Pertanto i valori destagionalizzati sono:

$$y_1^d = 16 - 13.5 = 2.5 \qquad y_3^d = -2 - (-5.5) = 3.5$$

$$y_2^d = 4 - 2.5 = 1.5 \qquad y_4^d = -9 - (-10.5) = 1.5$$

iv. Fornire delle previsioni per i valori assunti da y_t nei quattro trimestri successivi al periodo di osservazione:

$$\hat{y}_{61} = 13.5 + 0.5 \cdot 61 = 44 \qquad \hat{y}_{63} = -5.5 + 0.5 \cdot 63 = 26$$

$$\hat{y}_{62} = 2.5 + 0.5 \cdot 62 = 33.5 \qquad \hat{y}_{64} = -10.5.5 + 0.5 \cdot 64 = 21.5$$