Introduzione ai Modelli per Serie Storiche Economiche

Piano Lauree Scientifiche

Laboratorio TESTA-RDI

28 Febbraio 2014

Prof. Filippo DOMMA

(f.domma@unical.it)

Dipartimento di Economia, Statistica e Finanza – UniCal

Definizione. SERIE STATISTICA

Per serie statistica si intende un insieme di dati ordinati secondo un criterio qualitativo.

Quando il criterio ordinatore dei dati è il tempo, inteso come progressione cronologica, si ha una SERIE STORICA (o Temporale).

Definizione. SERIE STORICA (o Serie Temporale)

Una serie storica è una successione di dati numerici nella quale ogni dato è associato ad un particolare istante od intervallo temporale.

oppure

Definizione. SERIE STORICA (o Serie Temporale)

Per serie storica si intende un insieme di osservazioni ordinate secondo il tempo.

Serie Storica UNIVARITA:
Una s.s. è univariata se in ciascun punto o intervallo di tempo viene osservato un solo fenomeno (variabile). L'obiettivo principale dell'analisi è lo studio della **dinamica** del fenomeno in esame.

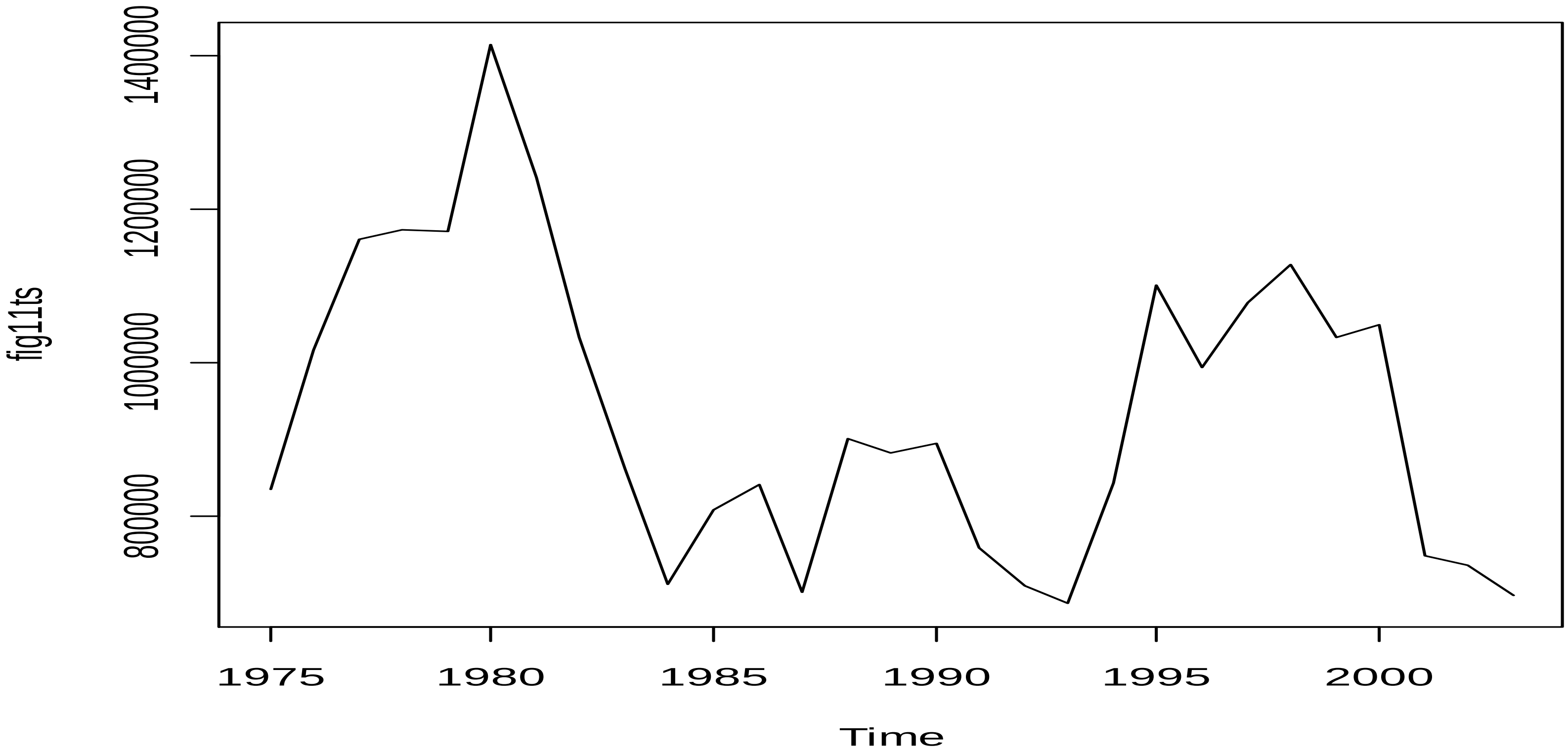
Serie Storica MULTIVARIATA:
Una s.s. è multivariata se in ciascun punto o intervallo di tempo si osservano due o più fenomeni (variabili) simultaneamente. In questo caso l'obiettivo principale dell'analisi è l'individuazione di eventuali **relazioni** esistenti tra i fenomeni.

ESEMPIO

Produzione italiana di motoveicoli dal 1975 al 2003
I dati sono archiviati nel file Excel *Figura_1-1.xlsx* contenuto nella cartella *PLS_Febbraio 2014*

833750
1016500
1160500
1172500
1170600
1412750
1239550
1032950
861050
711675
808250
840078
700046
899513
882492
894899
758826
708826
685310
842072
1101073
992433
1077705
1127788
1032137
1048184
749000
736500
697000

Fig. 1.1. Produzione italiana di motoveicoli, serie annuale



Osservazioni annuali rappresentati graficamente mediante un diagramma cartesiano in cui il **criterio ordinatore temporale** (anno) compare in ascissa e i valori osservati in ordinata.

In una s.s. è lecito presumere che vi sia **dipendenza** tra osservazioni successive;
la produzione di motoveicoli al tempo t è influenzata dalla produzione al tempo t-1 (e dai precedenti) ed influenza quella al tempo t+1 (ed i tempi successivi).

<u>Lo studio e la modellazione di tale dipendenza</u>, e la conseguente possibilità di sfruttarla ai fini previsivi, <u>rappresenta il cuore dell’analisi delle serie storiche</u>.

L'ambito di applicazione dell'analisi delle serie storiche è molto ampio.

Una possibile CLASSIFICAZIONE (non esaustiva) è la seguente:

Serie Storiche: **ECONOMICHE**, **FISICHE** e **DEMOGRAFICHE**

Serie GENERATE: DA CONTROLLO DI PROCESSO,
DA PROCESSI BINARI E
DA PROCESSI DI PUNTO

Serie SPAZIALI: si tratta di una serie particolare in cui il criterio ordinatore dei dati non è il tempo, ma lo spazio.

Serie Storica Economica

Investimenti fissi lordi a prezzi costanti (1995) registrati in Italia dal 1980 al 2003 (fonte: Istat)

Fig. 1.2. Investimenti fissi lordi, serie annuale (1980-2003)

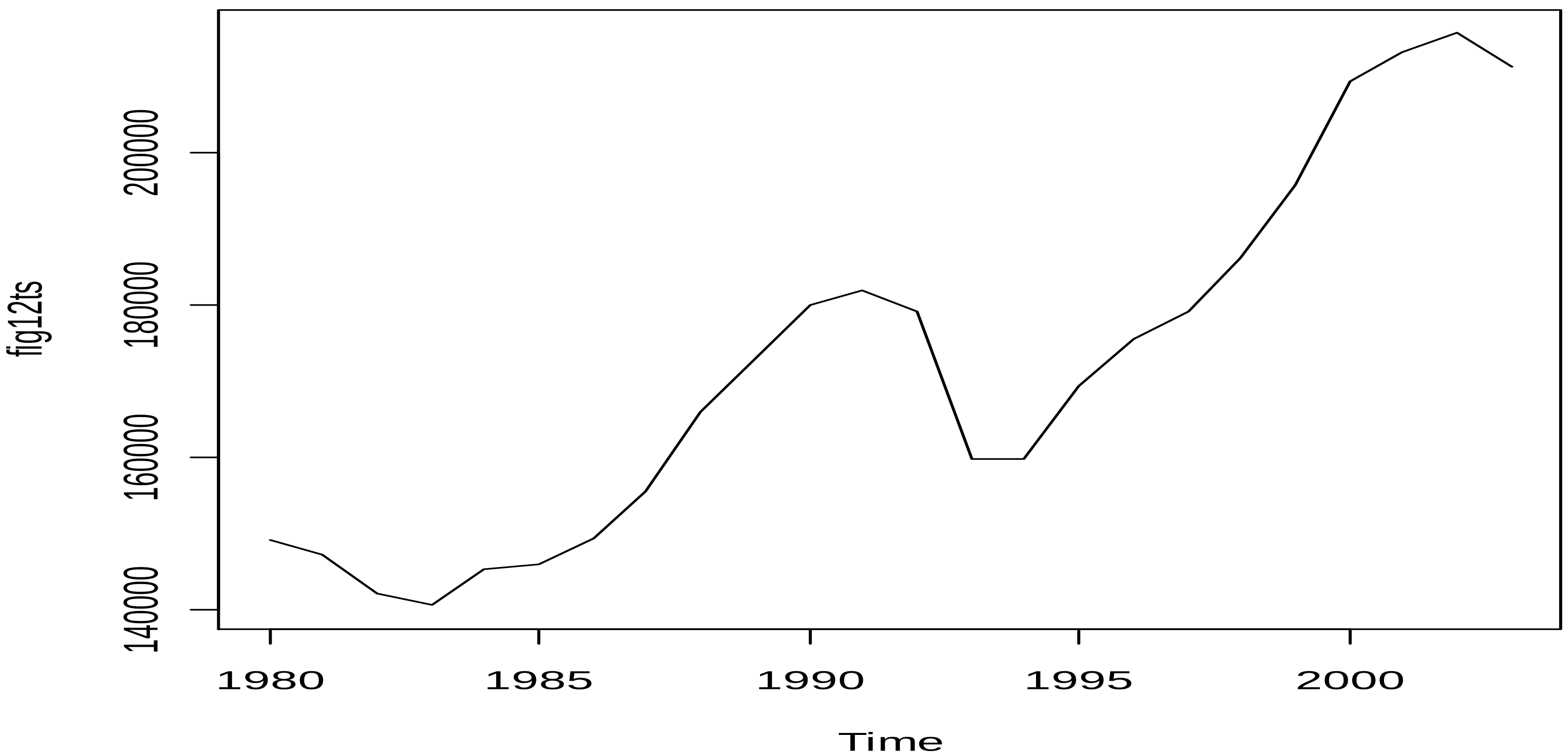


Fig. 1.3. Indice generale della produzione industriale, serie mensile (base 2000=100

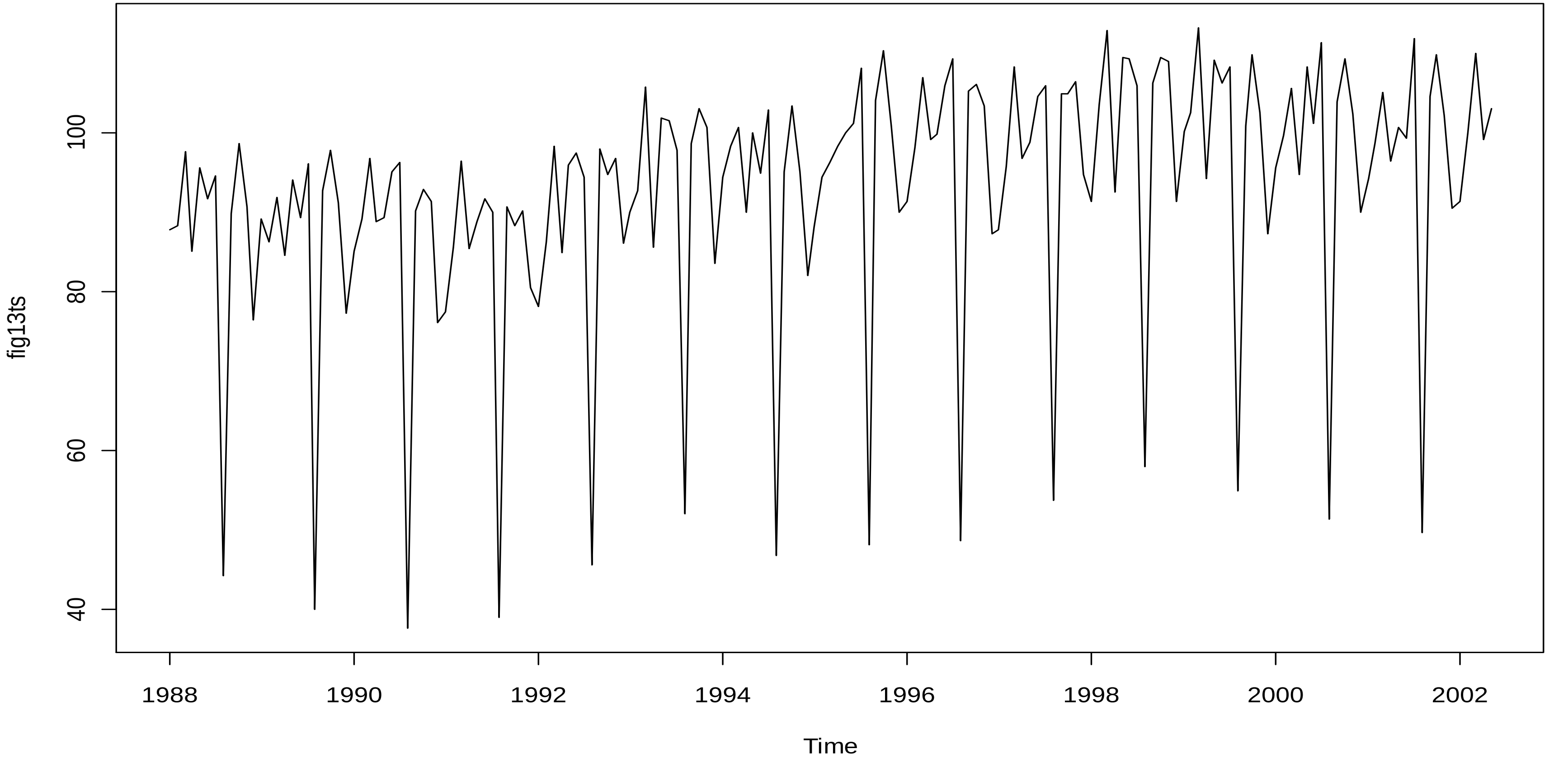
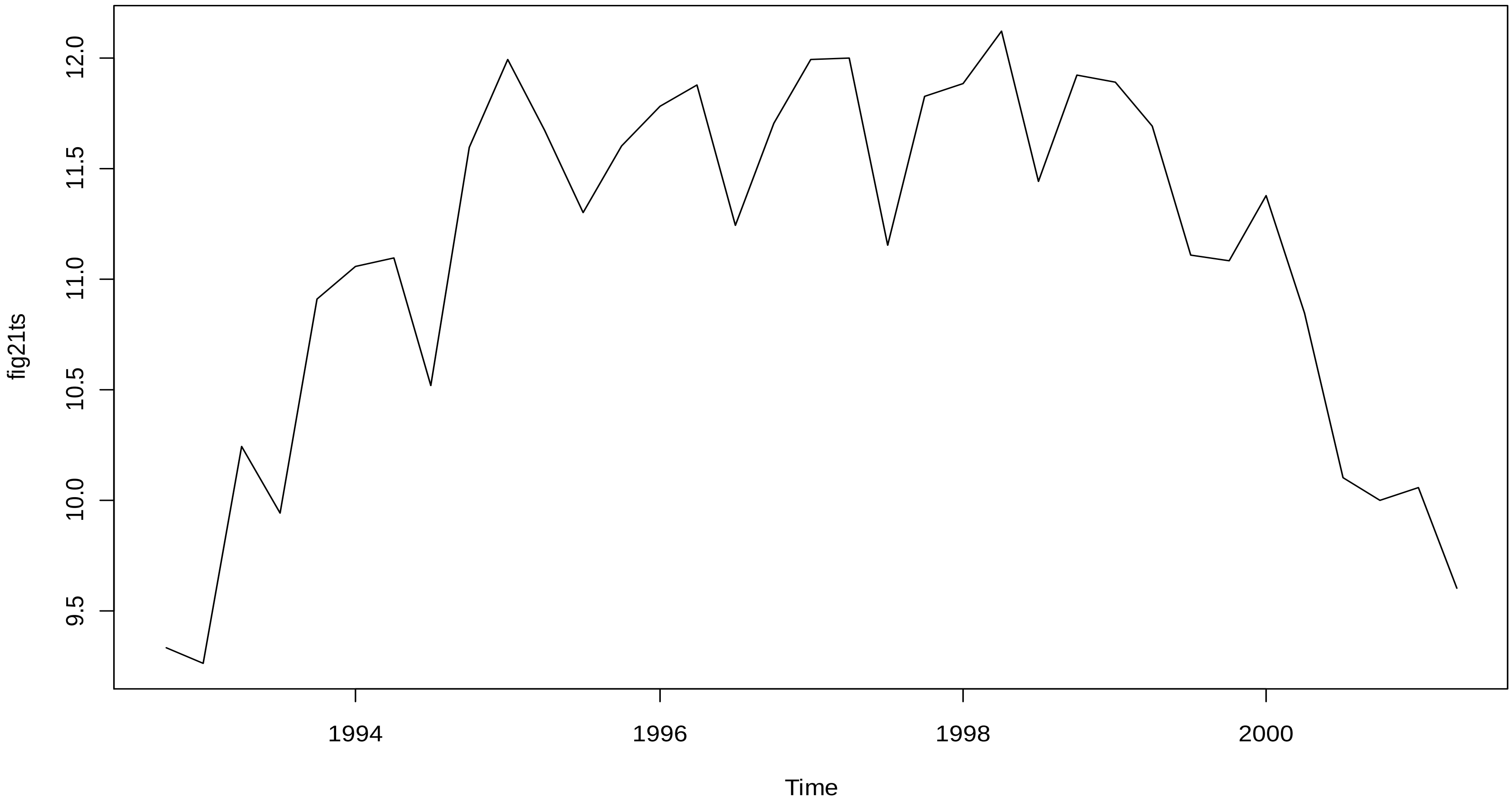


Fig. 2.1. Tasso di disoccupazione in Italia, serie trimestrale (1992.4-2001-2)



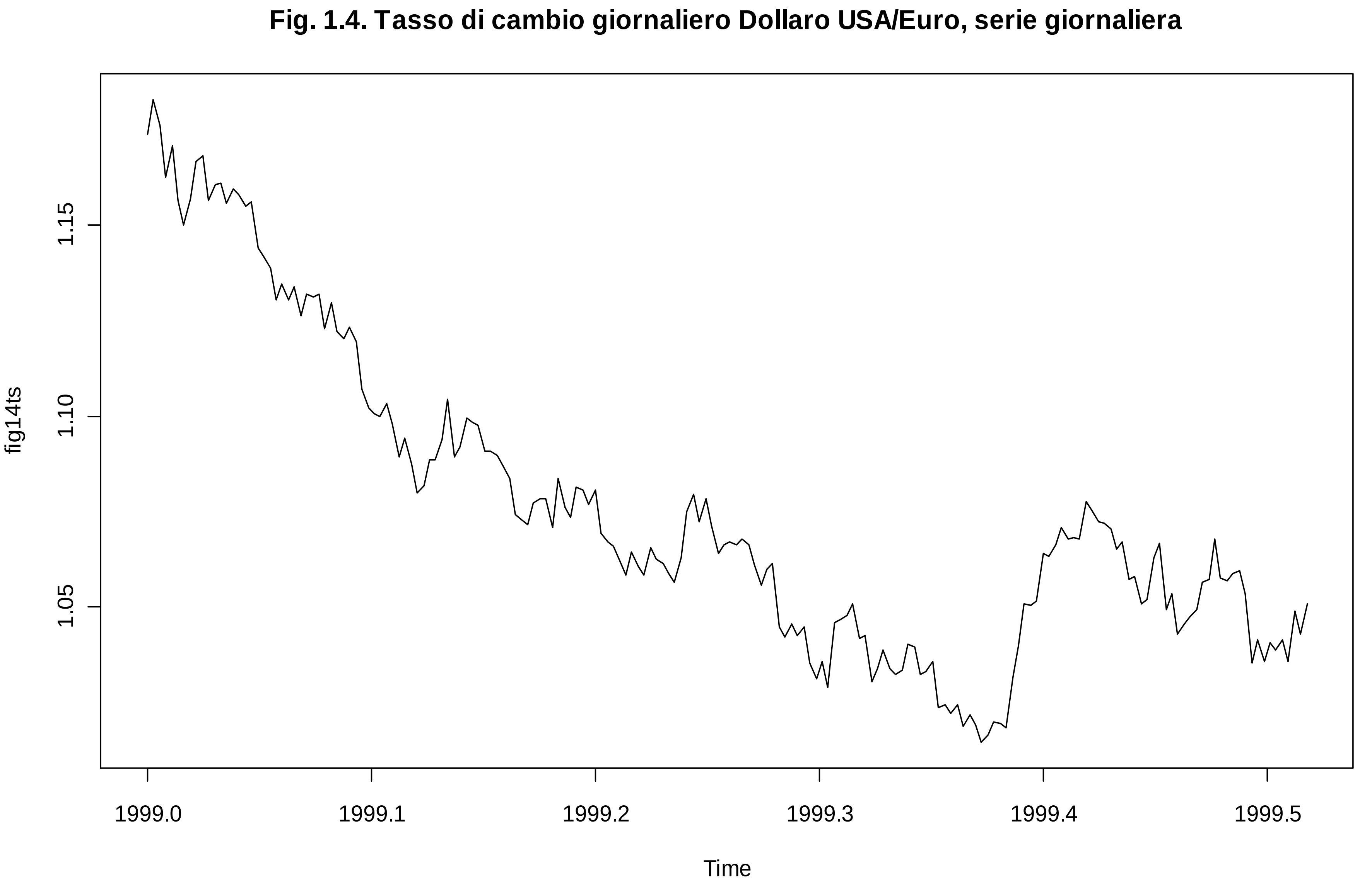
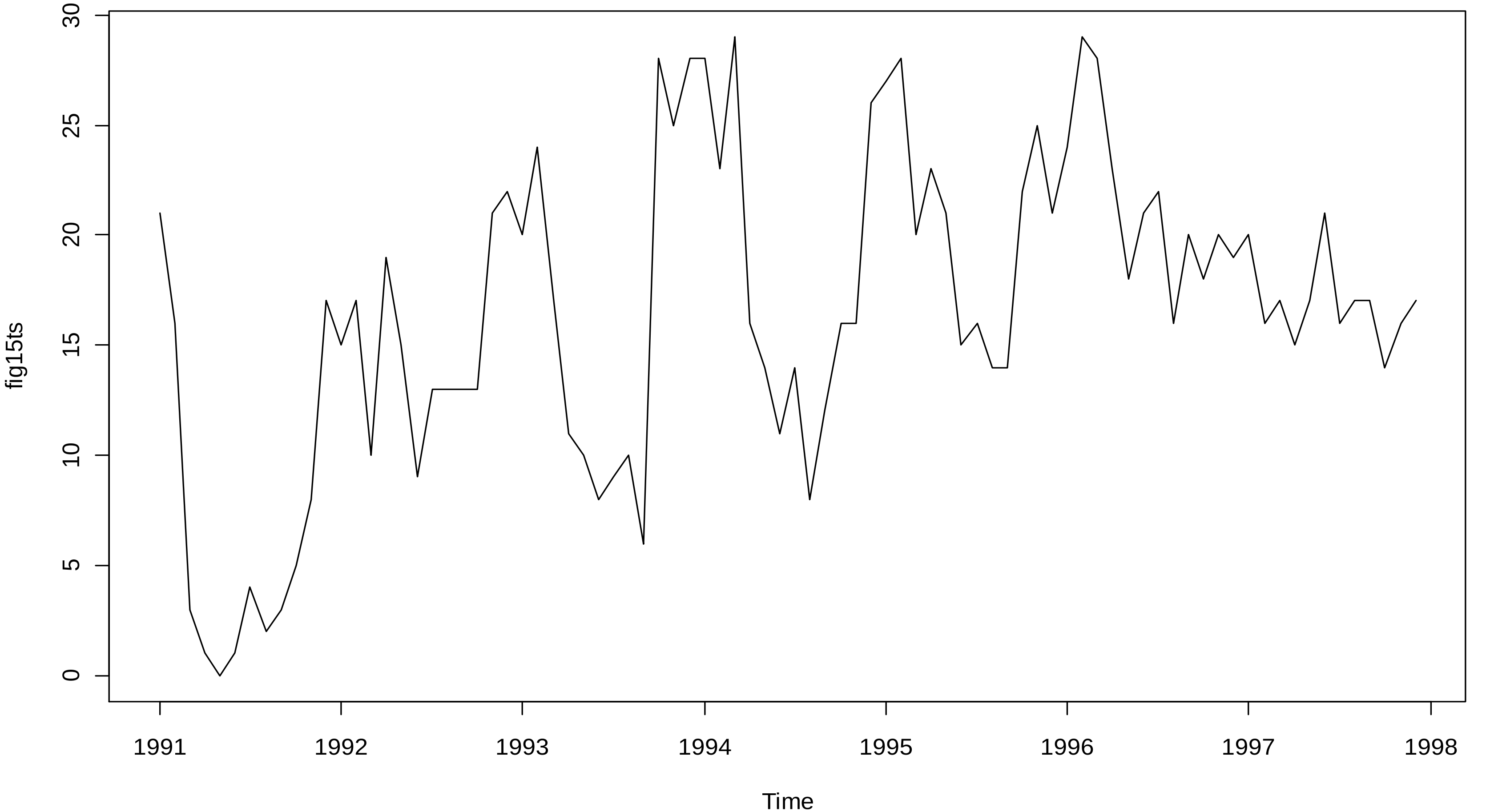


Fig. 1.4. Tasso di cambio giornaliero Dollaro USA/Euro, serie giornaliera
fig14ts
1.05
1.10
1.15
1999.0
1999.1
1999.2
1999.3
1999.4
1999.5
Time

Serie Storiche Fisiche

Fig. 1.5 Concentrazione mensile di anidrite solforosa - Padova (01.1991-12.1997



Serie Storiche Demografiche

Fig. 1.6. Quoziente di natalità italiano (%), serie mensile 01.1997-04.1999

Terminologia e Simbologia

Una serie storica viene indicata con

$$\{y_t\}_{t=1}^{n}$$

Il parametro temporale *t*, che definisce l'ordinamento temporale, appartiene ad un insieme parametrico *T*, che può essere continuo o discreto.

Nel primo caso, si ha una serie storica a parametro continuo (si evidenzia che il fenomeno osservato, *y*, può non essere continuo), nel secondo si ha una serie storica a parametro discreto (il fenomeno osservato, *y*, può non essere discreto).

In questa lezione, studieremo le <u>***serie storiche equispaziate a parametro discreto***</u>, cioè i casi in cui le sequenze di osservazioni del fenomeno oggetto di studio sono rilevate ad intervalli equidistanti con *t* appartenente all'insieme dei numeri interi relativi.
Per variabili di tipo economico gli intervalli temporali più utilizzati sono costituite da **giorni**, **settimane**, **mesi**, **trimestri** e **anni**.

Serie Storiche Deterministiche e Serie Storiche Stocastiche

Una s.s. $\{y_t\}_{t=1}^{n}$ è detta deterministica se esiste una funzione

$$\psi_t = \psi(t, y_t, y_{t-1}, y_{t-2}, \ldots\ldots)$$

tale che

$$E[y_t - \psi_t]^2 = 0 \qquad \text{per ogni t}$$

cioè se esiste una funzione della storia passata del fenomeno che determina y_t senza errore (dato che l'aspettativa è uguale a zero se e solo se $y_t = \psi_t$). Di conseguenza una serie deterministica può essere <u>prevista esattamente</u> sulla base della sua storia passata.

Se la funzione ψ è lineare allora la s.s. è detta <u>deterministica lineare</u>.

Esempio. La s.s.

$$y_t = a + bt$$ è lineare nei parametri.

Inoltre, dato che

$$y_t - y_{t-1} = a + bt - [a + b(t-1)] = b$$

$$y_{t-1} - y_{t-2} = a + b(t-1) - [a + b(t-2)] = b$$

si ha:

$$y_t - y_{t-1} = y_{t-1} - y_{t-2} \quad \Rightarrow$$

$$\Rightarrow \quad y_t = 2y_{t-1} - y_{t-2}$$

è deterministica (perché y_t dipende solo dalla sua storia passata (y_{t-1}, y_{t-2})

quindi, y_{t+1} sarà:

$$y_{t+1} = 2y_t - y_{t-1}$$

Esempio. La s.s. generata da un polinomio di 2° grado

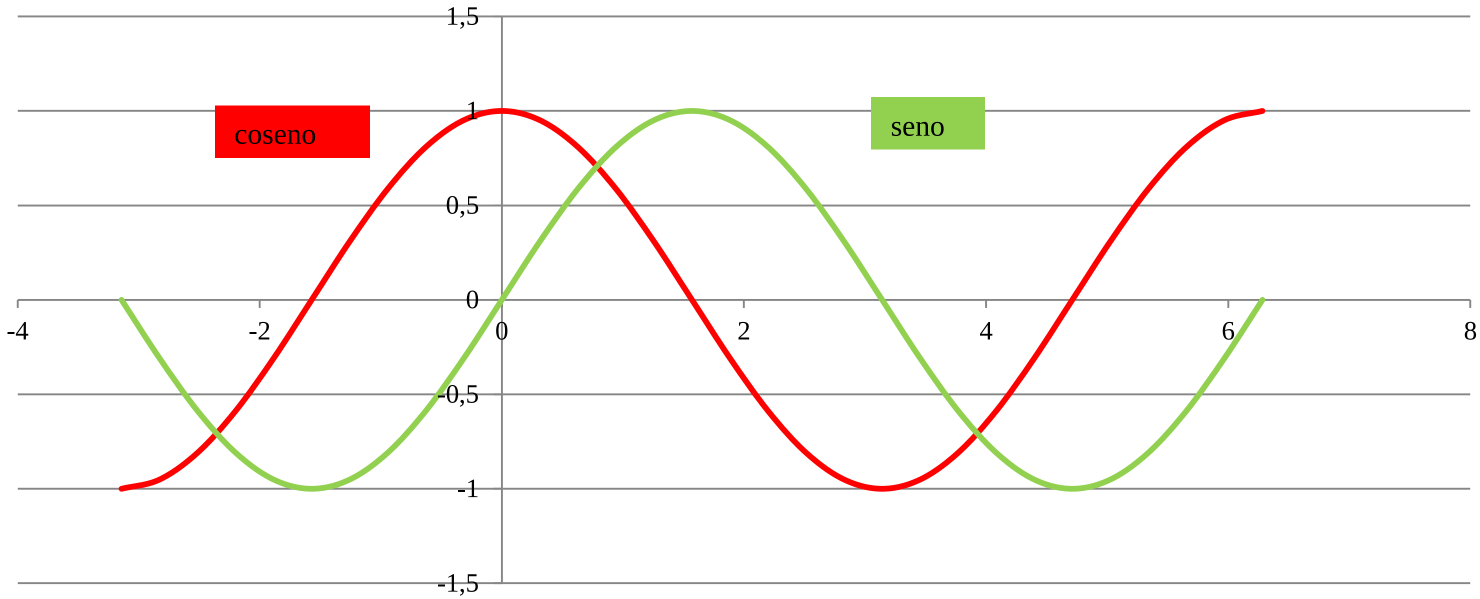
$$y_t = a_0 + a_1 t + a_2 t^2$$ è lineare nei parametri.

Verificare se è deterministica e calcolare quanto vale y_{t+1}

E’ utile ricordare la definizione di funzione periodica.

Definizione. Funzione Periodica.

Si dice che una funzione f(x) ha periodo P o che è periodica di periodo P se, per ogni x, f(x+P)=f(x) essendo P una costante positiva.



Approccio CLASSICO allo studio delle Serie Storiche

Un approccio sufficientemente generale, per descrivere il processo generatore dei dati di una serie storica

$$\left\{ y_t \right\}_{t=1}^{n}$$

relativa ad una variabile Y, è dato da

$$Y_t = f(t) + u_t \qquad \textbf{(2.3)}$$

Nella (2.3), la s.s. Y_t è suddivisa in due componenti:

-una sequenza completamente **deterministica**, f(t), che costituisce la parte sistematica della serie;

-una sequenza di variabili casuali, u_t , che rappresenta la parte **stocastica** della s.s.

Nell'ambito del cosiddetto approccio classico all'analisi delle s.s. si ipotizza che esiste una **legge di evoluzione temporale** del fenomeno rappresentata da f(t).

La componente casuale u_t rappresenta l'insieme delle circostanze, ciascuna di entità trascurabile, che non si possono considerare esplicitamente in Y_t. I residui di Y_t, non spiegati da f(t), vengono imputati al caso ed assimilati ad errori accidentali.

Da un punto di vista statistico, si ipotizza che u_t è generata da un processo *white noise*, cioè da una successione di v.c. iid, di media nulla e varianza costante; formalmente scriveremo:

$$u_t \sim WN(0, \sigma_u^2)$$

cioè

$$E(u_t) = 0 \qquad \forall\ t$$

$$V(u_t) = \sigma_u^2 \qquad \forall\ t$$

$$E(u_t u_r) = 0 \qquad \forall\ t \neq r$$

Per le serie storiche economiche generate dalla (2.3) si assume che la componente deterministica, f(t), sia il risultato dell'azione di tre componenti, direttamente non osservabili, ma chiaramente definibili sul piano concettuale: il **trend**, il **ciclo** e la **stagionalità** (per serie osservate con cadenza inferiore all'anno)

IL TREND: T_t

Il trend (o componente tendenziale) di una serie storica è la tendenza di fondo del fenomeno considerato, riferita ad un lungo periodo di tempo. Esso è determinato da fenomeni quali lo sviluppo economico e l'evoluzione strutturale del sistema economico, che per loro natura si manifestano con gradualità e lentamente nel tempo. Si assume che i valori del trend siano esprimibili tramite delle funzioni del tempo.

IL CICLO: C_t

Il ciclo (o componente congiunturale) è costituito dalle fluttuazioni attribuibili al succedersi di fasi ascendenti e fasi discendenti, generalmente collegate con le fasi di espansione e di contrazione del sistema economico. Tali fluttuazioni non sono di tipo regolare sia per la durata che per la loro ampiezza.

LA STAGIONALITA': S_t

La stagionalità (o componente stagionale) è costituita da movimenti del fenomeno nel corso dell'anno che tendono a ripetersi in modo pressoché analoga nel medesimo periodo di anni successivi.
Tali movimenti posso avere diverse cause: succedersi delle stagioni, variazioni climatiche, tradizioni ecc.

Modelli di combinazione delle componenti

a) Modello additivo: $Y_t = T_t + C_t + S_t + u_t$

b) Modello moltiplicativo: $Y_t = T_t C_t S_t u_t$

c) Modello misto: $Y_t = T_t C_t S_t + u_t$

Nel seguito, ci occuperemo della **stima** del TREND e della STAGIONALITA' sia utilizzando delle specificazioni analitiche (funzioni matematiche del tempo) per entrambi le componenti sia utilizzando dei metodi empirici (cioè senza l'ausilio di specificazioni analitiche).

ANALISI ESPLORATIVA DI UNA S.S.

Quando si analizza una variabile economica, spesso l'interesse non è incentrato sul valore assoluto della variabile, ma piuttosto sulle variazioni relative (tassi di crescita). In tal caso, si individua un tempo di riferimento (detto base), che viene mantenuto fisso, e si valuta la dinamica del fenomeno relativamente alla base. Formalmente, indicato con y_0 il valore del fenomeno al tempo base, il numero indice (a base fissa) di Y al tempo t è:

$$I_{0t} = \frac{y_t}{y_0} 100$$

La variazione relativa (percentuale) è data dal complemento a 100 del numero indice, cioè I_{0t} -100; così ad esempio, se I_{0t}=105.2 allora il valore di Y nel periodo *t* è superiore a quello del tempo base per una quota pari al 5.2% (cioè Y ha subito una variazione percentuale positiva -crescita percentuale- pari al 5.2% rispetto al tempo base).

Nel caso in cui si vuole confrontare il valore del fenomeno con quello del fenomeno precedente, si può calcolare il cosiddetto indice a base mobile (la variazione percentuale è sempre il complemento a 100)

$$I_{t-1,t} = \frac{y_t}{y_{t-1}} 100$$

E' utile osservare che se il fenomeno è stagionale allora ha senso calcolare i tassi di variazione rispetto allo stesso periodo dell'anno precedente al fine di ottenere una valutazione non influenzata dalla stagionalità. Ad esempio, dato che i consumi sono più elevati nel mese di Dicembre per effetto del Natale, si calcola

$$I_{t-12,t} = \frac{y_t}{y_{t-12}} 100$$

Richiamo: Modello Lineare $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

$$\mathrm{y}_{(n\times 1)} = \begin{pmatrix} y_1 \\ y_2 \\ .... \\ y_n \end{pmatrix} \qquad \mathbf{X}_{(n\times p)} = \begin{pmatrix} 1 & x_{11} & x_{12} & ... & x_{1k} \\ 1 & x_{21} & x_{22} & ... & x_{2k} \\ ... & ... & ... & ... & ... \\ ... & ... & ... & ... & ... \\ 1 & x_{n1} & x_{n2} & ... & x_{nk} \end{pmatrix} \qquad \beta_{(p\times 1)} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ .... \\ \beta_k \end{pmatrix} \qquad \boldsymbol{\varepsilon}_{(n\times 1)} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ .... \\ \varepsilon_n \end{pmatrix}$$

Stimatore ai minimi quadrati

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{X}^t\mathbf{y}$$

Indice di determinazione

$$R^2 = \frac{SQM}{SQT} = 1 - \frac{SQR}{SQT}$$

L’indice di determinazione corretto

Com’è noto, l’indice di determinazione R^2 aumenta (o resta costante) all’aumentare del numero di variabili esplicative inserite nel modello lineare, anche se l’aggiunta della nuova variabile esplicativa non ha senso logico nello spiegare la variabile dipendente.
Per ovviare a tale difetto si suole calcolare il cosiddetto <u>indice di determinazione corretto</u> dato da:

$$\overline{R}^2 = 1 - \frac{SQR/(n-p)}{SQT/(n-1)} = 1 - \left[\frac{n-1}{n-p}\left(1 - R^2\right) \right]$$

Si osserva che tale indice non può più essere interpretato come la frazione di variabilità totale spiegata dal modello stimato. Infatti, esso potrebbe assumere anche valori negativi.

Supponiamo che per la s.s. osservata y_t valga il modello

$$Y_t = f(t) + u_t \qquad \text{con} \qquad u_t \sim WN(0, \sigma_u^2) \qquad \textbf{(3.1)}$$

Supponiamo per il momento che la parte sistematica sia composta **solo** dal **TREND**.

Il trend è una componente che si caratterizza per un'evoluzione lenta e regolare nel corso del tempo. Per tale ragione esso è rappresentabile mediante una qualche **funzione del tempo** che deve, evidentemente, essere stimata.

Dopo aver effettuato le operazioni preliminari sulla s.s., l'analisi grafica della s.s. osservata può suggerire la funzione del tempo da utilizzare per stimare il trend, cioè può aiutarci a scegliere quella funzione che meglio descrive l'andamento di fondo della s.s. osservata. Usualmente sono utilizzati: i **polinomi**, le **curve esponenziali**, le funzioni **trigonometriche** e le **curve di crescita**.

Trend Polinomiale

Assumiamo che la funzione deterministica, f(t), sia un polinomio di grado *q* del tipo

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

Sostituendo f(t) nell'espressione della s.s., (3.1), si ha:

$$Y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + ... + \alpha_q t^q + u_t \qquad \textbf{(3.2)}$$

per t=1,2,…,n

E' evidente che la (3.2) non è altro che un modello di regressione lineare nei parametri, α_j , j=0,1,..,q stimabili con il **metodo dei minimi quadrati**. La (3.2) scritta in forma matriciale diventa:

$$\mathbf{y} = \mathbf{P}\boldsymbol{\alpha} + \mathbf{u}$$

dove

$$\mathbf{y} = \begin{pmatrix} y_1 \\ .. \\ y_t \\ .. \\ y_n \end{pmatrix} \qquad \mathbf{P} = \begin{pmatrix} 1 & 1 & 1^2 & ... & 1^q \\ 1 & 2 & 2^2 & ... & 2^q \\ .. & .. & .. & ... & .. \\ 1 & t & t^2 & ... & t^q \\ .. & .. & .. & ... & .. \\ 1 & n & n^2 & ... & n^q \end{pmatrix} \qquad \mathbf{u} = \begin{pmatrix} u_1 \\ .. \\ u_t \\ .. \\ u_n \end{pmatrix} \qquad \boldsymbol{\alpha} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ .. \\ .. \\ \alpha_q \end{pmatrix}$$

con p=q+1

Applicando il metodo dei minimi quadrati le stime dei parametri sono

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

Ottenute le stime ai m.q. possiamo stimare la varianza tramite i residui stimati

$$\hat{\sigma}_u^2 = \frac{1}{n-p}\hat{\mathbf{u}}_t'\hat{\mathbf{u}}_t = \frac{1}{n-p}\sum_{t=1}^{n}\hat{u}_t^2$$

dove $\hat{u}_t = \hat{y}_t - y_t$ con $\hat{\mathbf{y}}_t = \mathbf{P}\hat{\boldsymbol{\alpha}}$

e sotto l'ipotesi di normalità degli errori costruire intervalli di confidenza e test d'ipotesi.

L'ordine *q* del polinomio dipende dal comportamento di fondo della s.s. osservata. Casi particolari

$$Y_t = \alpha_0 + u_t$$ Trend costante

$$Y_t = \alpha_0 + \alpha_1 t + u_t$$ Trend lineare

$$Y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + u_t$$ Trend parabolico

Se q è molto elevato allora
- Migliora l'accostamento delle ordinate stimate a quelle osservate
- Aumenta la complessità e diminuiscono i gradi di libertà

Criteri di scelta dell'ordine del polinomio
- delle differenze successive
- confronto tra $\overline{R}^2$

Criterio del confronto tra $\overline{R}^2$

Un criterio ragionevole per la scelta dell'ordine del polinomio si fonda sulla versione

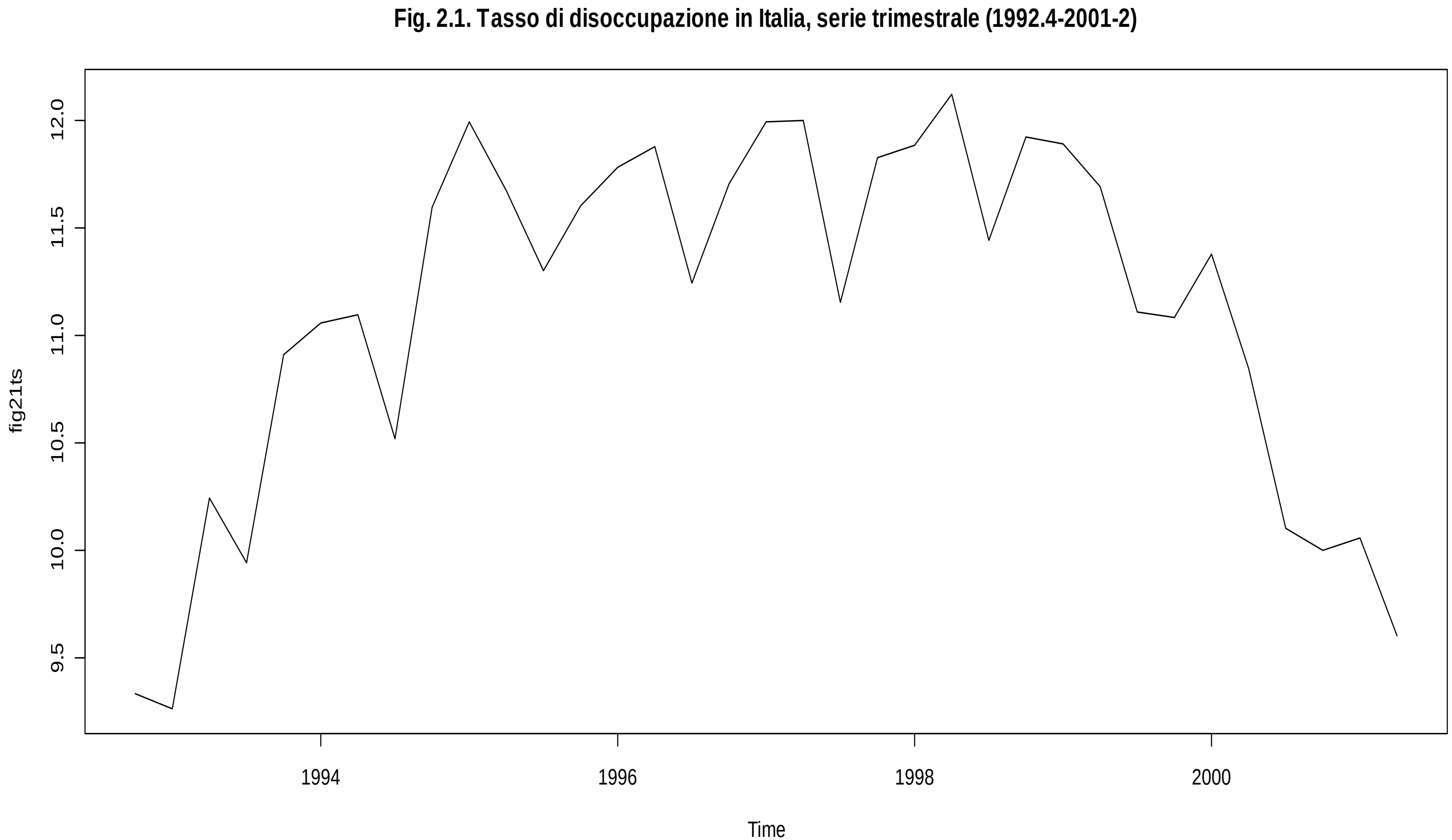
corretta del coefficiente di determinazione. In particolare, indicato con $\overline{R}_q^2$

il coefficiente di determinazione corretto relativo alla stima di un polinomio di ordine q, se

$$\overline{R}_q^2 \geq \overline{R}_{q+1}^2$$

allora si sceglie un polinomio di grado q, altrimenti si prosegue nella ricerca per valori di q superiori. Evidentemente, se è soddisfatta la disuguaglianza sopra riportata vuol dire che la componente t^{q+1} non ha fornito un contributo significativo all'adattamento del modello stimato ai dati.

Esempio. Stima del TREND: Tasso di disoccupazione in Italia – Dati trimestrali 1992.4 – 2001.4



Cosa suggerisce il grafico ?

Per determinare il **grado del polinomio** utilizziamo il metodo del confronto tra indici di determinazione corretti.

Bisogna costruire dei polinomi in t, per t=1,2,…,n, dove n è la dimensione della serie.

Vedi file Figura_2-1.xlsx

	Polinomio di ordine 1
R al quadrato corretto	-0,031236425

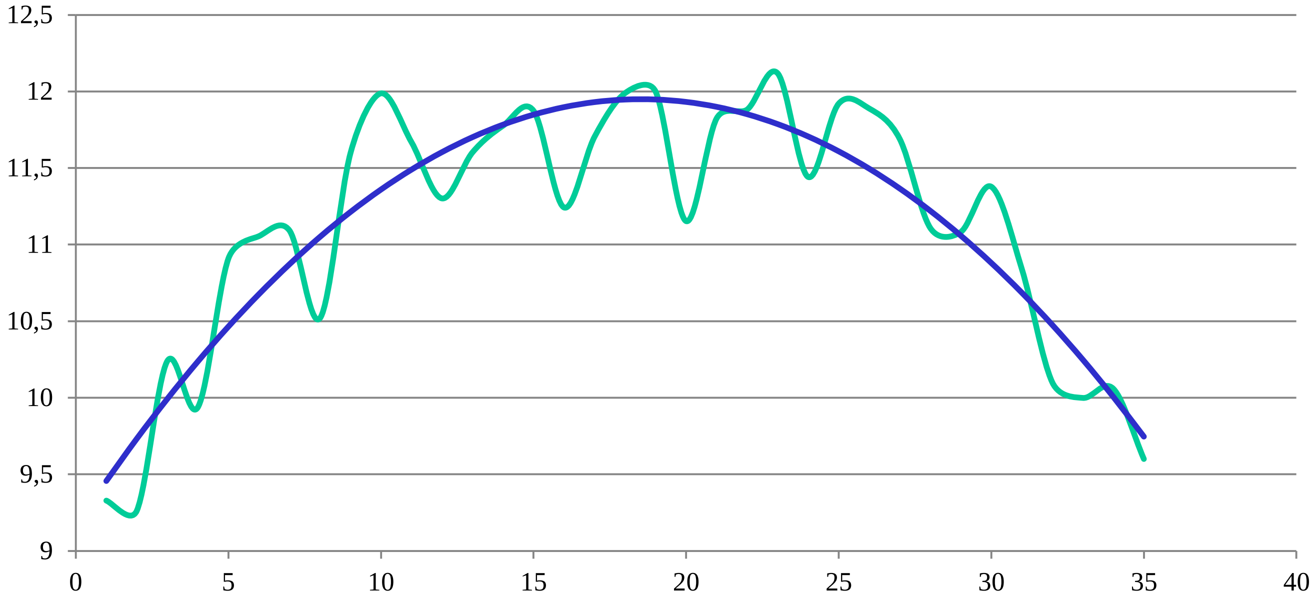
	Polinomio di ordine 2
R al quadrato corretto	0,822872202

	Polinomio di ordine 3
R al quadrato corretto	0,817212117

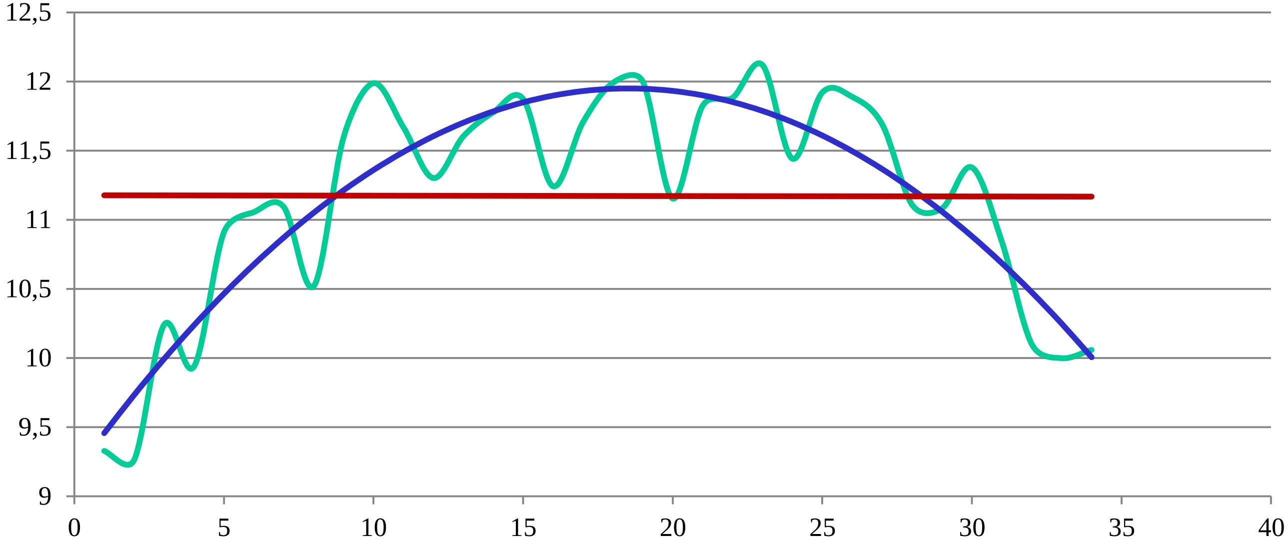
Il modello stimato (trend stimato) è:

$$\hat{Y}_t = 9.16371 + 0.30076t - 0.008118t^2 \qquad t=1,..,n$$

Stima del TREND – Polinomio di ordine 2.



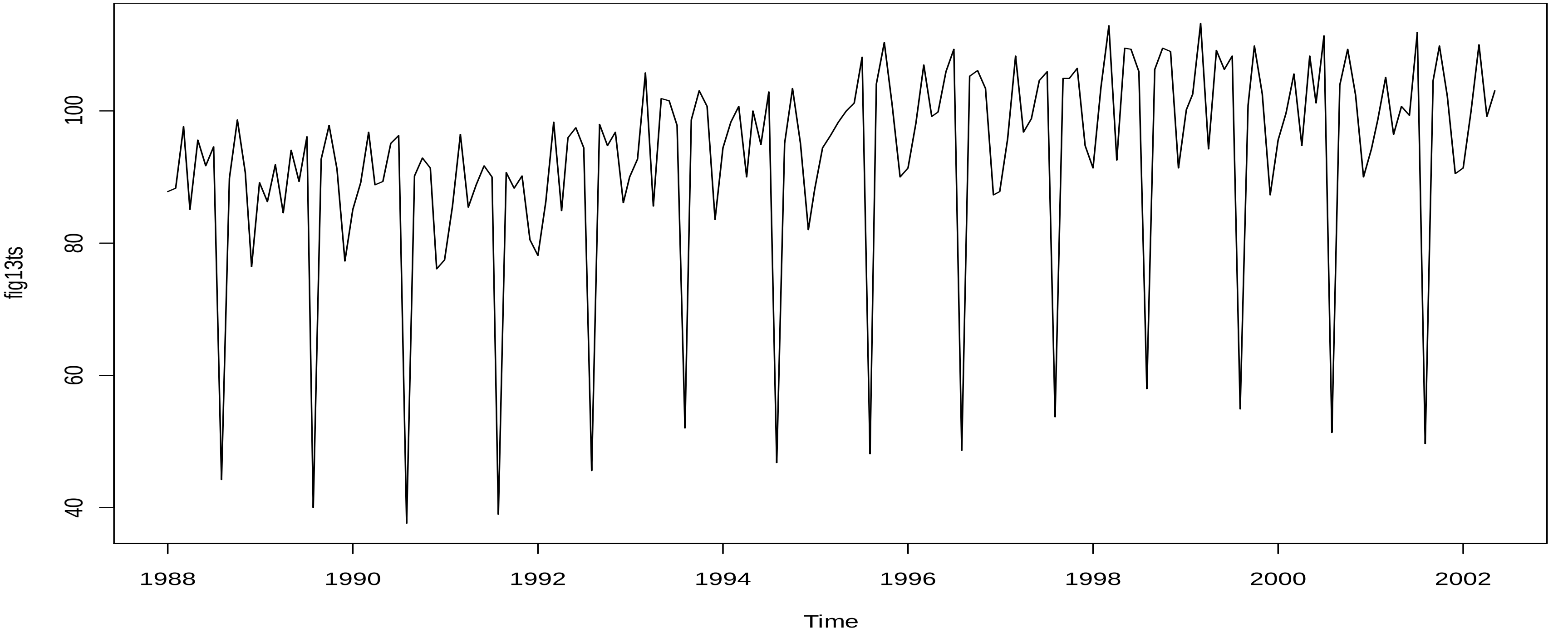
Curiosità: la linea rossa stima un TREND Lineare !



Stima della componente stagionale

La **stagionalità** (o componente stagionale) è costituita da movimenti del fenomeno nel corso dell'anno che tendono a ripetersi in modo pressoché analoga nel medesimo periodo di anni successivi. Tali movimenti posso avere diverse cause: succedersi delle stagioni, variazioni climatiche, tradizioni ecc.

Fig. 1.3. Indice generale della produzione industriale, serie mensile (base 2000=100



La componente stagionale di una s.s. può essere rappresentata da una **funzione periodica** g(t). Ricordiamo che le funzioni periodiche sono quelle funzioni il cui valore all'istante *t* si riproduce esattamente ad intervalli costanti, la cui lunghezza *s* costituisce il periodo, ossia

$$g(t) = g(t+s) = g(t+2s) = g(t+3s) = \ldots\ldots$$

osserviamo che per serie trimestrali s=4, per serie mensili s=12 …..

Esempio. La funzione seno, g(t)=sin(t), è periodica di periodo s=2π, cioè sin(t)=sin(t+2π)=sin(t+4π)=...; la funzione tangente è periodica di periodo π, cioè tan(t)=tan(t+π)=tan(t+2π)=….

Esempio. La s.s. **trimestrale** è osservata 4-volte in un anno, s=4. Così, se la componente stagionale è osservata nel secondo trimestre, avremo:

$$g(2) = g(2+4) = g(2+8) = g(2+12) = \ldots\ldots$$

Compon. Stag. osservata nel 2° trimestre

Compon. Stag. osservata nel 2° trimestre dell'anno successivo, perché si ripete dopo s=4 periodi (trimestri)

Compon. Stag. osservata nel 2° trimestre del secondo anno successivo, perché si ripete dopo s=8 periodi (trimestri)

Il modello di **regressione** può essere utilizzato per stimare la componente stagionale, tramite due metodi:

-mediante le variabili ausiliarie dicotomiche (*dummy*)

-mediante combinazioni di funzioni trigonometriche

Supponiamo per il momento che nella s.s. sia presente **solo** la componente stagionale S_t cioè, si assume che

$$Y_t = S_t + u_t$$

La s.s. Y_t è specificata dalla componente stagionale S_t e dalla componente erratica u_t

Metodo delle variabili *dummy*

Supponiamo che la funzione periodica sia rappresentabile tramite

$$g(t)=\sum_{j=1}^{s}\gamma_j d_{j,t}=\gamma_1 d_{1,t}+\gamma_2 d_{2,t}+\ldots+\gamma_s d_{s,t} \qquad \text{t=1,2,…,n}$$

dove $d_{j,t}$ è una variabile *dummy* indicatrice del periodo nell'anno t, cioè

$$d_{j,t}=\begin{cases}1 & \text{nel periodo j dell'anno t}\\ 0 & \text{altrimenti}\end{cases}$$

Facendo riferimento ai dati <u>trimestrali </u>(s=4), si ha:

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

Ad esempio, $d_{4,16}$ indica la variabile dummy in corrispondenza della 16-esima osservazione della s.s. per il 4° trimestre, assume valore 1 se la 16-esima osservazione cade nel 4° trimestre, altrimenti vale zero. È evidente che se la 16-esima osservazione ricade effettivamente nel 4° trimestre allora $g(16)=\gamma_4$; analogamente, se la 20-esima osservazione cade effettivamente nel 4° trimestre si ha: $g(20)=\gamma_4$ e così anche $g(24)=\gamma_4$, $g(28)=\gamma_4$ ….. Ciò evidenzia che la funzione sopra definita è periodica perché $g(4)=g(4+4)=g(4+2*4)=g(4+3*4)=g(4+4*4)=g(4+4*5)=\gamma_4$

Vediamo che valori assumono variabili *dummy* per dati trimestrali (s=4)

Osservazione	t	d1,t	d2,t	d3,t	d4,t
1a	I-2000	1	0	0	0
2a	II-2000	0	1	0	0
3a	III-2000	0	0	1	0
4a	IV-2000	0	0	0	1
5a	I-2001	1	0	0	0
6a	II-2001	0	1	0	0
7a	III-2001	0	0	1	0
8a	IV-2001	0	0	0	1
9a	I-2002	1	0	0	0
….	….	…	….	…	…

Specificazione

Assumiamo che il processo generatore della s.s. sia dato da:

$$Y_t = S_t + u_t \qquad (3.1)$$

in cui S_t =g(t) è la componente stagionale e u_t è la componente erratica. Vogliamo stimare la componente stagionale con il metodo dei minimi quadrati.

Ad esempio, nel caso di dati **trimestrali** osservati per N anni completi (N=n/s) ed in assenza della componente trend, il modello di regressione associato con la (3.1) è:

$$\mathbf{y} = \mathbf{D}\boldsymbol{\gamma} + \mathbf{u} \qquad (3.2)$$

Poiché i dati sono **trimestrali** la componente stagionale, per un generico *t*, è data da:

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

e, quindi, la matrice **D** dei regressori contenenti le variabili dummy del modello (3.2) è:

$$\mathbf{D} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{array}{l} \left.\begin{array}{c} \\ \\ \\ \\ \end{array}\right\} t=1 \\ \left.\begin{array}{c} \\ \\ \\ \\ \end{array}\right\} t=2 \\ \begin{array}{c} \\ \\ \end{array} \\ \left.\begin{array}{c} \\ \\ \\ \\ \end{array}\right\} t=N \end{array}$$

$$\boldsymbol{\gamma} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{pmatrix}$$

Coefficienti di regressione rappresentano il livello che mediamente il fenomeno assume in corrispondenza di ciascuno degli s periodi infrannuali (trimestri, mesi, ..)

N = numero di anni completi

Ritorniamo al nostro modello di regressione per la componente stagionale nel caso

$$\mathbf{y} = \mathbf{D}\boldsymbol{\gamma} + \mathbf{u}$$

Utilizzando il metodo dei minimi quadrati, la stima del vettore dei coefficienti di regressione è:

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

I coefficienti $\hat{\gamma}^*_j$ per j=1,2,…,s, sono chiamati **coefficienti grezzi di stagionalità**

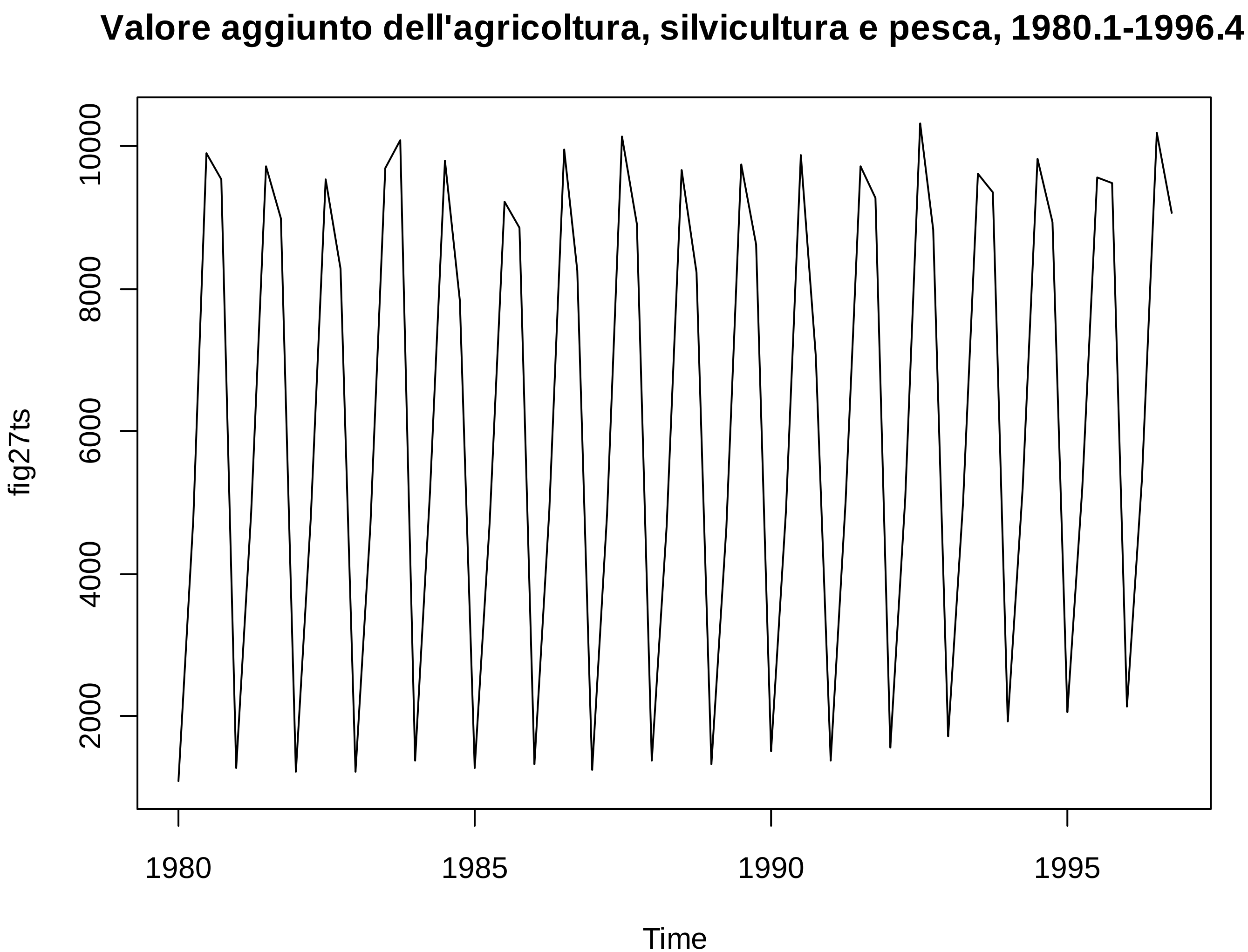
Una prima stima della componente stagionale è

$$\hat{g}(t) = \sum_{j=1}^{s} \hat{\gamma}^*_j d_{j,t} \qquad \text{t=1,2,…,n}$$

in notazione matriciale

$$\hat{\mathbf{g}} = \mathbf{D}\hat{\boldsymbol{\gamma}}^*$$

Il grafico riporta la serie **trimestrale** del valore aggiunto ai prezzi di mercato dell'agricoltura, silvicoltura e pesca (1980.1-1995.4). La serie mostra un chiaro andamento stagionale e sembra evolvere attorno ad un **trend costante**. Si può ipotizzare un modello additivo con la <u>sola componente stagionale</u>.

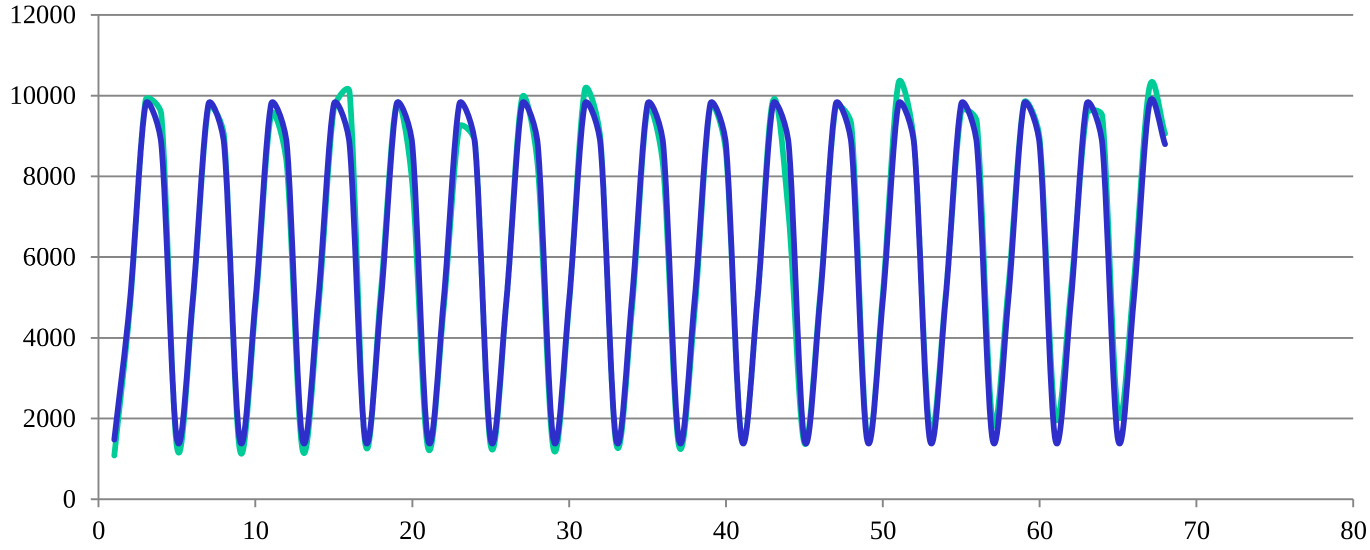


Esempio: file Excel *Figura_2-7.xlsx*

Modello stimato: componente stagionale

$$\hat{y}_t = 1475.94 d_{1,t} + 4921.29 d_{2,t} + 9787.588 d_{3,t} + 8795.88 d_{4,t}$$

Confronto tra dati osservati e componente stagionale stimata



Per **<u>destagionalizzazione</u>** si intende quella operazione finalizzata ad <u>eliminare</u> la componente stagionale in una serie storica. Di conseguenza, per **s.s. destagionalizzata** si intende la serie <u>depurata</u> dalla componente stagionale.
Dato il modello $Y_t = S_t + u_t$, la s.s. destagionalizzata con i coefficienti grezzi **coincide con la serie dei residui** (stimati)

$$y_t^{*d} = y_t - \hat{g}(t)$$

In forma matriciale

$$\mathbf{y}^{*d} = \mathbf{y} - \hat{\mathbf{g}}$$

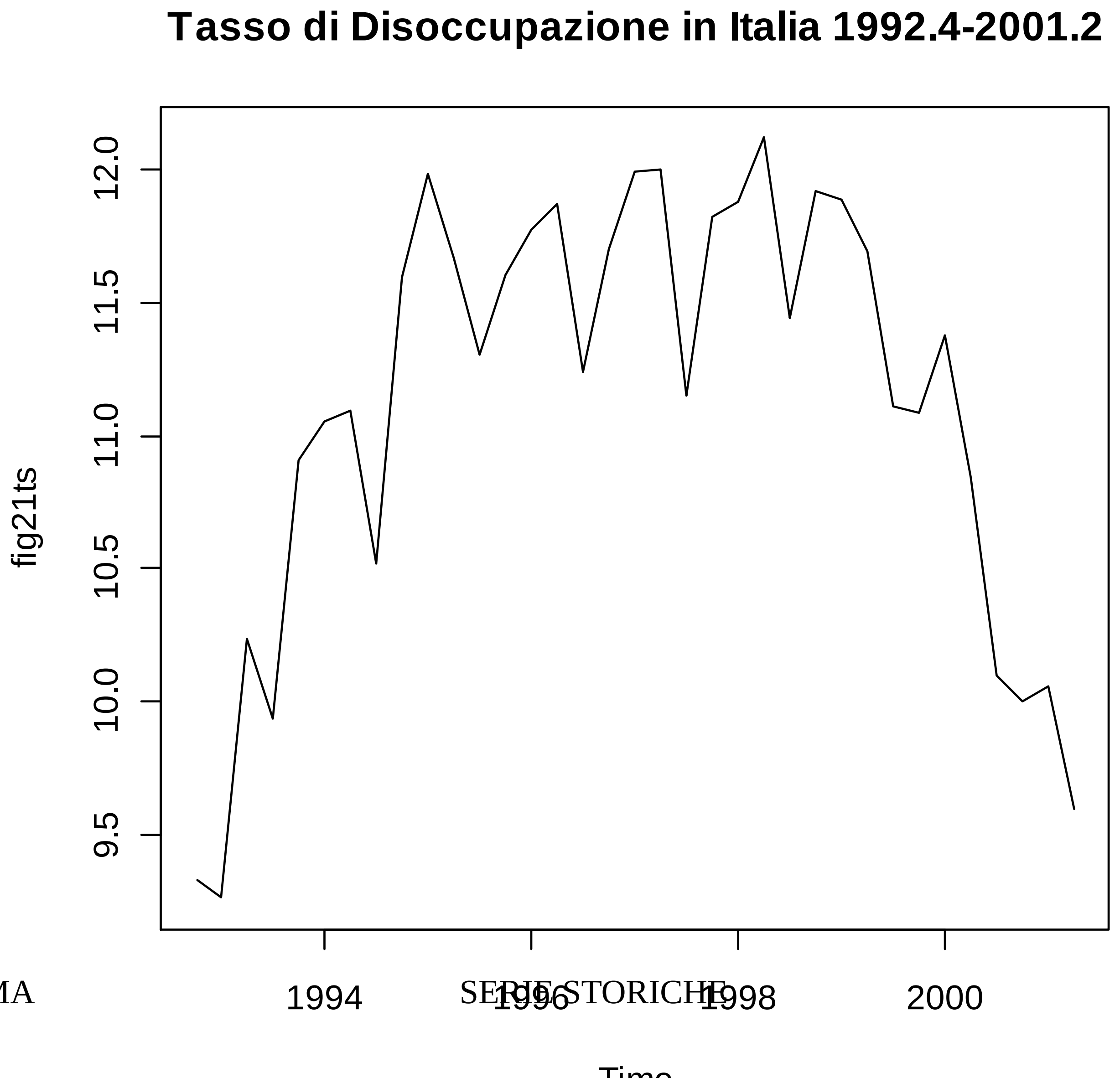
Stima di un TREND **polinomiale** e di una componente STAGIONALE **con le dummy**

Volendo stimare contemporaneamente il trend e la componente stagionale, possiamo combinare, ad esempio, un trend polinomiale e una componente stagionale espressa tramite le variabili Dummy; per dati trimestrali abbiamo

$$y_t = \alpha_0 + \alpha_1 t + .. + \alpha_q t^q + \beta_1 d_{1,t} + \beta_2 d_{2,t} + \beta_3 d_{3,t} + \beta_4 d_{4,t} + \varepsilon_t$$

otteniamo così un modello lineare nei parametri α_j per j=0,1,…,k, β_h per h=1,2,3,4.
Per la stima di detti parametri possiamo utilizzare il metodo dei minimi quadrati.

Esempio. Tasso di disoccupazione in Italia
Abbiamo già stimato un TREND quadratico per questa serie trimestrale
Proviamo ad aggiungere una componente stagionale utilizzando le variabili dummy.


Tasso di Disoccupazione in Italia 1992.4-2001.2
fig21ts
12.0
11.5
11.0
10.5
10.0
9.5
1994
1996
1998
2000
Time

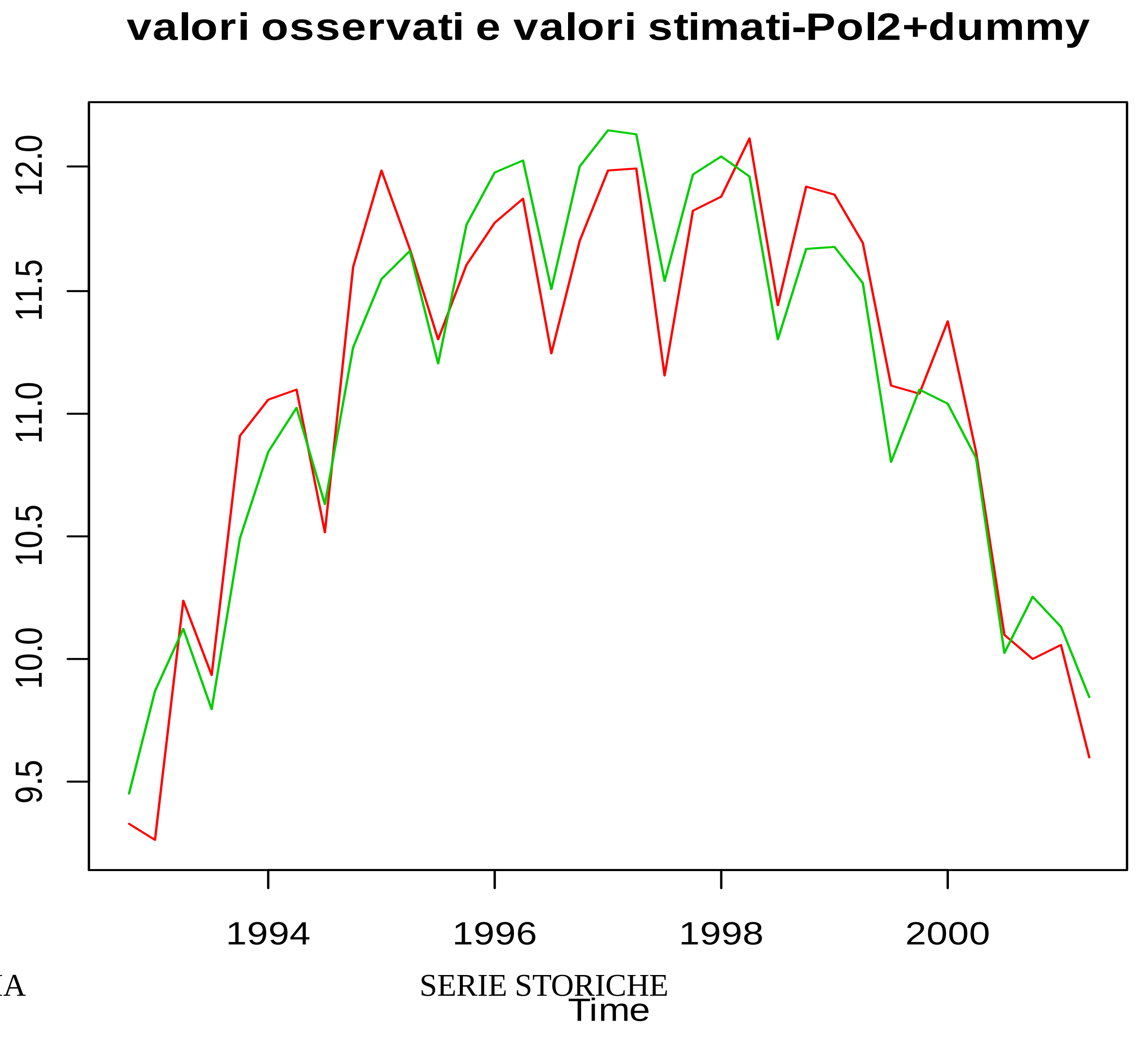
Vogliamo stimare il seguente modello

$$y_t = \alpha_1 t + \alpha_2 t^2 + \beta_1 d_{1,t} + \beta_2 d_{2,t} + \beta_3 d_{3,t} + \beta_4 d_{4,t} + \varepsilon_t$$

Stima: file Excel *Figura_2-1.xlsx*

Modello stimato: Trend polinoniale + Componente Stagionale

$$\hat{y}_t = 0.31013t - 0.00839t^2 + 9.282d_{1,t} + 9.265d_{2,t} + 8.689d_{3,t} + 9.153d_{4,t}$$


valori osservati e valori stimati-Pol2+dummy
9.5
10.0
10.5
11.0
11.5
12.0
1994
1996
1998
2000
Time