Università degli Studi della Calabria
Arcavacata di Rende (Cosenza)

Dipartimento di Economia e Statistica

La metodologia Box-Jenkins
(quello che ancora ne rimane)

Prof. Agostino Tarsitano



Tutto scorre, Tutto torna

Anno Accademico 2012-2013

Perché ti sei iscritto all'università?
- Perché?! Chiesero la stessa cosa a un tale che saltò a piedi nudi su un cesto di fichi d'India.
- E lui cosa rispose?
- Che da principio, gli era sembrata una buona idea.

Variante dal film "I magnifici Sette di John Sturges

Indice

1 Processi stocastici **1**

1.1 Introduzione . . . 1

1.1.1 Serie storiche, memoria e metodologia Box-Jenkins . . . 2

1.1.2 Il processo stocastico . . . 5

1.1.3 La serie storica . . . 7

1.1.4 Processo stocastico stazionario . . . 9

1.1.5 Il correlogramma . . . 11

1.1.6 Processo stocastico stazionario in senso debole . . . 14

1.1.7 Processo *white noise* . . . 17

1.1.8 Processi gaussiani . . . 18

1.1.9 Ergodicità . . . 19

1.1.10 Ergodicità e autocorrelazioni . . . 22

1.1.11 Autocorrelazioni parziali . . . 23

2 Analisi preliminari **27**

2.1 Correzioni di calendario . . . 28

2.1.1 Effetto giorni lavorativi . . . 29

2.1.2 Feste mobili (Effetto pasqua) . . . 31

2.2 Trasformazione dei dati . . . 32

2.2.1 Trasformazioni contro l'eteroschedasticità . . . 34

2.3 Valori anomali nella serie storica . . . 40

2.3.1 Riflessioni sui valori anomali . . . 44

2.3.2 Individuazione dei valori anomali . . . 47

2.3.3 Trattamento dei valori anomali . . . 53

2.4 Valori mancanti . . . 55

2.4.1 Ragioni dei valori mancanti . . . 55

2.4.2 Mancate risposte parziali . . . 57

2.4.3 Dati mancanti perchè impossibili . . . 58
2.4.4 Dati volutamente omessi . . . 58
2.5 Strutture possibili dei dati mancanti . . . 59
2.5.1 Dati mancanti completamente a caso (MCAR) . . . 63
2.5.2 Dati mancanti a caso (missing at random, MAR) . . . 64
2.5.3 Dati mancanti non a caso (MNAR) . . . 65
2.6 Valori mancanti in una serie storica . . . 67
2.6.1 Imputazione dei dati mancanti . . . 69
2.7 Differenziazione di una serie storica . . . 75

3 Processi stocastici **79**
3.1 Introduzione . . . 79
3.1.1 Funzione di auto-covarianza e di auto-correlazione per processi lineari . . . 83
3.1.2 Stazionarietà ed invertibilità dei processi lineari . . . 84
3.2 Processi auto-regressivi . . . 87
3.2.1 Processo AR(1) . . . 89
3.2.2 Processo AR(2) . . . 91
3.3 Processi media mobile . . . 93
3.3.1 Processo MA(1) . . . 95
3.3.2 Processo MA(2) . . . 97
3.4 Processi ARMA . . . 99
3.4.1 Auto-correlazioni semplici e parziali . . . 101
3.4.2 Processo ARMA$(1,1)$ . . . 103
3.4.3 Processi ARIMA . . . 104

4 Modelli SARIMA **109**
4.1 Introduzione . . . 109
4.2 Arima Stagionali (SARIMA) . . . 112
4.2.1 Differenziazione stagionale . . . 114
4.2.2 Esempi di modelli SARIMA . . . 117
4.2.3 Modelli moltiplicativi ARIMA e SARIMA . . . 122
4.3 Esempio applicativo . . . 127
4.3.1 Studio dei correlogrammi . . . 130

5 Stima, Identificazione e validazione degli ARIMA **131**
5.1 Stima . . . 131
5.1.1 Stima di massima verosimiglianza . . . 131
5.1.2 Stima di massima verosimiglianza degli ARIMA . . . 133
5.2 Adeguatezza del modello . . . 136
5.2.1 Misure di adattamento . . . 136
5.2.2 Misure di adattamento e parsimonia . . . 137

5.2.3 Indici di erraticità . . . 139
5.2.4 Indici di gaussianità . . . 142
5.3 Identificazione . . . 146
5.3.1 Identificazione automatica . . . 148
5.4 Possibilità e limiti delle previsioni . . . 154
5.4.1 Le previsioni nelle scienze sociali . . . 159
5.4.2 Previsioni con la metodologia BJ . . . 160
5.4.3 Attendibilità delle previsioni (*cross-validation*) . . . 164

6 RegArima **167**
6.1 Processo Trend+ARMA . . . 167

Elenco delle figure

1.1 Esempi di serie storica 2
1.2 Idea della serie storica 4
1.3 Grafico da un canale di ECG 5
1.4 Processi stocastico e sue realizzazioni 6
1.5 Molteplicità delle serie storiche originatesi in un processo 8
1.6 Clusters di serie storiche 8
1.7 Serie storica non stazionaria 9
1.8 Andamento delle autocovarianze 11
1.9 Esempi di correlogrammi 12
1.10 Regione di esistenza di una valida funzione di autocorrelazione per $n = 2$ 14
1.11 Serie storica non stazionaria in media a causa di un trend 15
1.12 Precautionary demand component of the real price of oil 16
1.13 White noise 18
1.14 Esempio di ergodicità 20
1.15 Wind records are classified as stationary (left) or nonstationary (right) 21
1.16 Serie storia breve e correlogramma 22
1.17 Esempi di correlazioni parziali 26

2.1 Cambio dell'unità di tempo 28
2.2 Tipologie delle coppie di mesi e giornate lavorative 30
2.3 Giornate lavorative negli anni 30
2.4 Le date della Pasqua 31
2.5 Simulazione di quattro serie storiche con marginali diverse 33
2.6 Frazionamento in quattro sotto serie 34
2.7 Verifica grafica della stazionarietà in media ed in varianza 35
2.8 Caso in cui la trasformazione potrebbe risultare utile 36

2.9 Effetto della scala logaritmica . . . 36
2.10 Esempio di trasformazione . . . 37
2.11 Effetto della trasformazione . . . 38
2.12 Alcuni esempi della distribuzione Pearson III . . . 39
2.13 Anomalie nelle serie storiche brevi . . . 41
2.14 Valori anomali in fenomeni evolutivi . . . 41
2.15 Valori anomali in fenomeni evolutivi . . . 42
2.16 Tracciato di un elettrocardogramma . . . 42
2.17 Valori anomali in fenomeni a coda pesante . . . 43
2.18 Tracciato di un elettrocardogramma . . . 44
2.19 Scoperte nuove o peculiarità insospettate . . . 46
2.20 Boxplot per individuare gli *outliers* . . . 48
2.21 Rappresentazione delle soglie di anomalia . . . 48
2.22 Soglie di anomalia e serie storiche . . . 49
2.23 Calcolo delle soglie di anomalia . . . 49
2.24 Serie D di Box-Jenkins. Due valori sospetti con $\alpha = 10\%$ . . . 50
2.25 Metodo della mediana per la mondatura (*cleaning*) dei dati . . . 51
2.26 Valore anomalo contestuale . . . 52
2.27 Calcolo delle soglie di anomalia . . . 53
2.28 Puzzle . . . 60
2.29 Esempio di dati mancanti in un data set . . . 61
2.30 Esempio di una matrice di mancanze . . . 61
2.31 Tipologie di dati mancanti . . . 63
2.32 MCAR . . . 64
2.33 MAR per due variabili . . . 65
2.34 Esempio di meccanismo MNAR informativo . . . 67
2.35 Esempio di serie storica con valori mancanti . . . 68
2.36 Imputazione con il metodo della media . . . 70
2.37 Serie storica con valori imputati . . . 71
2.38 Imputazione con la mediana di Hodges-Lehmann . . . 72
2.39 Esempio del metodo di imputazione tabellare . . . 73
2.40 Esempio del metodo di imputazione con interpolazione . . . 73
2.41 Esempio del metodo di imputazione stagionale . . . 73
2.42 Imputazione "home made . . . 74
2.43 Esempio di trend quadratico . . . 75
2.44 Effetto del filtro delle differenze . . . 76
2.45 Differenziazione regolare della serie storica . . . 77
2.46 Esempio di eliminazione di un trend quadratico . . . 78

3.1 Processi stocastici e processi lineari . . . 80
3.2 Effetto del polinomio caratteristico . . . 80

3.3 Filtro agente nella direzione opposta . . . 83
3.4 Andamento dei pesi . . . 84
3.5 Pesi convergenti per un processo stocastico lineare . . . 85
3.6 Processi auto-regressivi del primo ordine . . . 90
3.7 Processi quasi *white noise* e processo *random walk* . . . 90
3.8 Esempi di AR(2) . . . 92
3.9 Altri esempi di AR(2) . . . 92
3.10 Esempi di processi MA del primo ordine . . . 95
3.11 MA con radici unitarie e quindi non invertibili . . . 96
3.12 Esempio di MA(2) . . . 97
3.13 Esempio di MA(2) . . . 98
3.14 Esempio di MA(2) . . . 98
3.15 ARMA$(1,1)$, $\phi_1 = 0.5, \theta_1 = 0.5$ . . . 104
3.16 ARMA$(1,1)$, $\phi_1 = -0.7, \theta_1 = -0.5$ . . . 105
3.17 ARMA$(1,1)$, $\phi_1 = -0.7, \theta_1 = 0.5$ . . . 105
3.18 ARMA$(1,1)$, $\phi_1 = 0.7, \theta_1 = -0.5$ . . . 106
3.19 Number of Hospitals within the health care system . . . 108

4.1 Organizzazione dei dati stagionali o periodici . . . 111
4.2 Esempio di serie storica stagionale sul consumo di energia . . . 112
4.3 Serie storica quadrimestrale sul consumo di metano nel Regno Unito . 113
4.4 La duplice natura delle differenze non necessarie. . . . 115
4.5 Organizzazione dei dati stagionali o periodici . . . 116
4.6 Esempio di processo con schema IMA$(0,0,1)_{12}$ . . . 118
4.7 Esempio di processo con schema ARI$(1,0,0)_{12}$ . . . 119
4.8 Esempio di spettro empirico di un ARI$(1,0,0)_{12}$ . . . 120
4.9 Esempio simulato di processo con schema ARIMA$(1,0,1)_{12}$ . . . 121
4.10 Esempio di spettro empirico di un ARIMA$(1,0,1)_{12}$ . . . 121
4.11 Accostamento di due componenti separate . . . 122
4.12 Monthly motor vehicles engines - Canada: 1976-1991 . . . 127
4.13 Monthly motor vehicles engines - Canada. Analisi del trend . . . 127
4.14 Analisi spettrale per lo studio della stagionalità . . . 128
4.15 Valutazione dell'ordine delle differenze regolari . . . 129
4.16 Valutazione dell'ordine delle differenze stagionali . . . 129
4.17 Correlogramma della serie *motor vehicles engines* . . . 130

5.1 *Test* dei segni . . . 140
5.2 *Test* dei segni . . . 141
5.3 *Test* dei segni in caso di indipendenza . . . 142
5.4 Grafici per valutare la gaussianità dei residui . . . 145
5.5 Schema tradizionale della procedura BJ . . . 146
5.6 Identificazione automatica della serie *motor vehicles engines* . . . 150

5.7 Stima del modello identificato 150
5.8 Identificazione automatica della serie *motor vehicles engines* 150
5.9 Identificazione automatica e stima di un AR ad alto ritardo 151
5.10 Analisi della correlazione nei resiui per ARIMA$(3,1,2)\times(0,1,1)_{12}$. 151
5.11 Analisi della casualità dei residui per ARIMA$(3,1,2)\times(0,1,1)_{12}$ . . 152
5.12 Analisi della correlazione nei residui per ARIMA$(25,0,0)$ 152
5.13 Ulteriore analisi della correlazione nei residui per ARIMA$(25,0,0)$ . . 153
5.14 Analisi della casualità nei residui per ARIMA$(25,0,0)$ 153
5.15 Schema generale delle previsioni 156
5.16 Esempio di ottimismo seriale (Irvine Fisher, Yale) 157
5.17 Guida spericolata 158
5.18 Previsioni con il modello ARIMA$(3,1,2)\times(0,1,1)_{12}$ 164
5.19 Previsioni con il modello AR(13) 166

6.1 ARMA$(1,1)\,,\phi_1=0.7,\theta_1=-0.5$ 168

Elenco delle tabelle

2.1 *Esempi per il test di Jarque-Bera* 33
2.2 Due serie storiche brevi 40
2.3 Test t dei due campioni (test di Welch) 56
2.4 Esempi di valori manifestatamente illogici 59

3.1 Scelta dell'ordine delle differenze 107

5.1 *Confronto della GOF per due modelli ARIMA* 166

CAPITOLO 1

Processi stocastici

1.1 Introduzione

Il primo passo nello studio di dati temporali è comprendere il legame tra la variabile ed il tempo: i valori si realizzano col trascorrere degli intervalli di tempo e queste sono perciò le unità su cui si osserva la variabile di interesse. La serie storica è una sequenza di valori per una variabile rappresentativa di un fenomeno la cui unità di rilevazione sia disposta ad intervalli successivi di tempo. Il valore della variabile si forma dopo che l'unità di tempo è trascorsa o ad un istante nel suo interno. In simboli, abbiamo

$$\{W_t | t = 1, 2, \cdots, n\} \qquad n \text{può essere infinito} \tag{1.1}$$

Il pedice t" evoca il trascorrere del tempo. Ecco alcuni esempi:

Fenomeno	Variabile	Unità
Portata di un fiume	F_t	secondi
Entrate orarie in un POS	E_t	ora
Chiusura indice di borsa	C_t	giorno comm.
Passaggi settimanali alla frontiera	P_t	settimana
Variazione mensile indice dei prezzi	V_t	mese
Indice trimestrale della produz. Indus.	Y_t	trimestre
Popolazione residente nei comuni al 31/12	R_t	anno

Una serie storica è detta continua o discreta in relazione al tipo di fenomeno in cui è acquisita. In pratica, il tempo t sarà sempre discreta, anche se gli intervalli di tempo non sono necessariamente regolari.

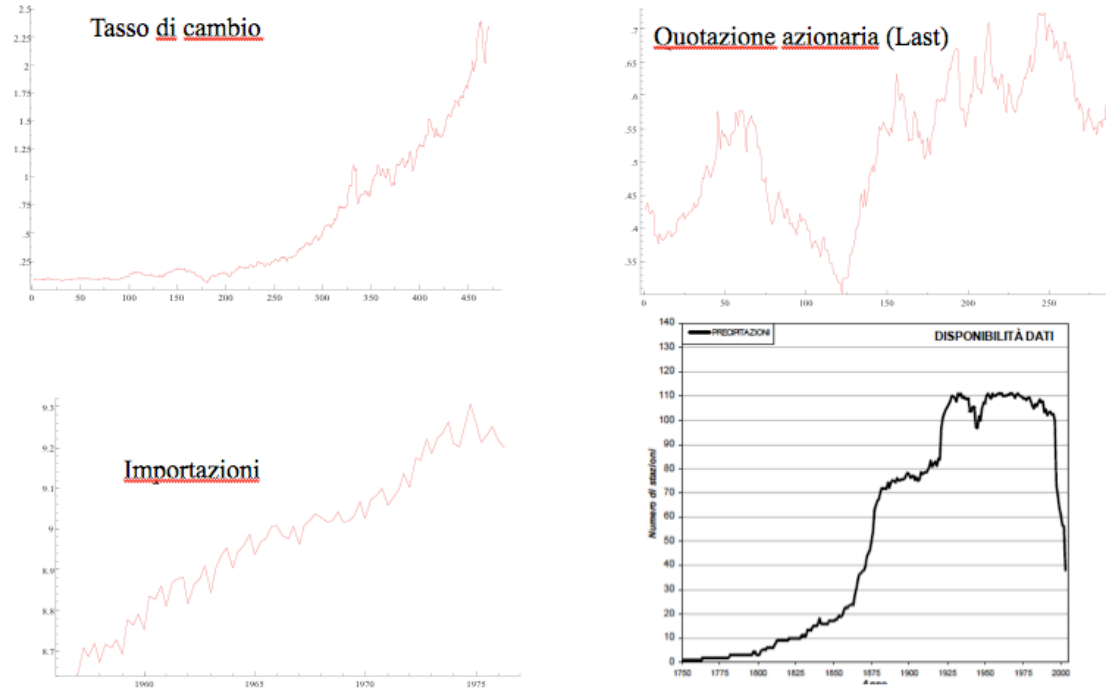


Figura 1.1: Esempi di serie storica

In base alla definizione si arguisce che per le serie storiche:

1. I dati si dispongono ad intervalli più o meno regolari di tempo.
2. Il significato del dato si arricchisce del fatto di essere osservato prima e/o dopo di un altro.
3. L'ordinamento dei dati non può essere alterato.

1.1.1 Serie storiche, memoria e metodologia Box-Jenkins

In molte occasioni i valori rilevati costituiscono, da un punto di vista statistico, le osservazioni di un campione casuale di variabili casuali indipendenti. L'analisi delle serie storiche è impostata sulla natura della dipendenza tra i membri della sequenza. Questa angolatura modifica la trattazione di molti fenomeni come distribuiti casualmente nel tempo intorno ad una media sostanzialente stabile. Parzen (1983) sostiene -con buone ragioni- che lo studio delle serie storiche o serie temporali deve essere imperniato sul concetto di memoria o persistenza o isteresi. Questa idea si può sintetizzare nella seguente nominazione:

- Serie storica con nessuna memoria (*white noise*).
- Serie storica con memoria corta (processo ergodico, ma non *white noise*).

- Serie storica con memoria media (presenza di trend, stagionalità, ciclo, non stazionario).
- Serie storica con memoria lunga (processo non ergodico).

Il concetto di ergodicità, meglio analizzato più in avanti, è fondamentale per l'analisi delle serie storiche e dei modelli statistici dinamici che nascono intorno alle serie storiche. In estrema sintesi l'ergodicità implica che ogni serie storica è un esempio di tutte le serie storiche dello stesso genere che si sarebbero potute osservare, ma che sono rimaste nelle potenzialità inespresse del fenomeno esaminato. Se si accerta che la serie storica è del primo tipo cioè a memoria breve, l'analisi può proseguire ad esempio per individuare il modello di distribuzione più adatto per descriverne i valori, ma rimane poco utile a fini previsivi. Se la serie risulta a memoria breve può essere analizzata con la modellistica introdotta nel nostro corso (approccio Box-Jenkins). Sulla serie a memoria media si può intervenire per trasformarla in una serie a memoria breve mentre questo non è possibile per le serie a memoria lunga (fenomeni a coda pesante) che impongono una modellistica più sofisticata. Se un evento particolare, casuale o presumibilmente tale, può cambiare il corso di una serie storica, allora si tratta di una serie non ergodica. Le serie storiche a memoria lunga implicano una forma di dipendenza tra le osservazioni di epoche diverse che persiste anche quando le epoche sono molto distanti tra di loro sull'asse temporale.

La metodologia Box-Jenkins (BJ) ha il merito di scaturire da problemi di grande interesse pratico. Fu Yule che nel 1927 introdusse la possibilità di spiegare il valore presente di un fenomeno che si evolve nel tempo ipotizzandone la dipendenza lineare da alcuni o tutti i suoi valori passati (fattore autoregressivo o AR). A questo si aggiunse il contributo di Slutsky che ne 1937 presentò gli schemi media mobile (MA) nel tentativo di spiegare l'errore di previsione attuale in termini degli errori trascorsi legati anch'essi da una qualche forma di dipendenza lineare. Comunque toccò a Wold nel 1938 realizzare l'accostamento tra le due parti autoregressive e le parti media mobile per formulare gli schemi ARMA.

Secondo Wold, una serie storica può essere espressa con la relazione

$$W_t = \mu + \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} + a_t - \theta_1 a_{t-1} - \theta_1 2 a_{t-2} - \cdots - \theta_q a_{t-q} \quad (1.2)$$

Il parametro μ rappresenta il livello medio della serie storica W_t. Se si specificano correttamente l'ordine p della componente AR e l'ordine q della componente MA, si possono modellare moltissime serie storiche osservate nella realtà ai fini della previsione, del monitoraggio e della valutazione dell'impatto di un certo intervento.

La semplicità dei modelli ARMA è però solo apparente. Infatti, le prime applicazioni concrete delle serie storiche avvennero nella seconda metà degli anni '60 del secolo scorso allorché si resero disponibili e si diffusero dei computer adeguati ai calcoli complessi nascosti sotto la superficie degli ARMA (autoregressivi-media mobile). Il punto di svolta fu però la pubblicazione del testo di Box e Jenkins (1970)

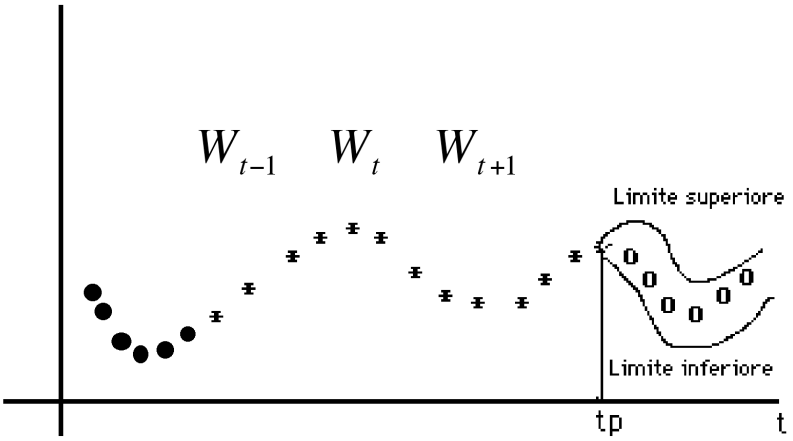


Figura 1.2: Idea della serie storica

che si riuscì ad organizzare in modo coerente e funzionale il materiale che si era accumulato nei decenni precedenti. Il mondo della finanza, delle telecomunicazioni, delle aziende, degli enti gestori di servizi (*web*, carte di credito, banche, sicurezza) sono solo alcuni dei contesti interessati dalla analisi delle serie storiche particolarmente per i suoi aspetti finalizzati ad ottenere previsioni agevoli ed attendibili.

Predictions are difficult. Especially about the future. Mark Twain

L'obiettivo di questa relazione è di mettere gli studenti in grado di utilizzare i modelli Box-Jenkins per l'analisi delle serie storiche. Si tratta di modelli il cui obiettivo primario è di determinare, sulla base dei valori già noti di un fenomeno che si evolve nel tempo, i suoi valori futuri o passati. In questa versione, non si mettono in relazione fenomeni diversi, non si considerano variabili endogene o esogene; neanche il tempo- nella forma di semplice variabile esplicativa appare in primo piano. Le tecniche introdotte si limitano ad estendere al di fuori del periodo per cui si hanno dati disponibili, le linee di tendenza mostrate nel periodo su cui si hanno invece i dati.

Il successo dell'approccio BJ deriva dalla capacità di generare modelli in grado di mimare il comportamento nel tempo di diversi fenomeni soggetti a variazioni stocastiche che presentano tuttavia genesi molto diverse. A questo risultato si arriva con pochi parametri e pochi accorgimenti e con delle linee guida efficaci, addirittura automatizzabili. E' chiaro che da questi modelli non possiamo pretendere chiavi di lettura o schemi interpretativi sofisticati: sono troppo elementari. Risultano invece validi per fornire previsioni, specie quelle a breve termine, tanto da superare in qualità di risultati, varie metodologie econometriche in vigore all'epoca della loro prima apparizione

1.1.2 Il processo stocastico

Consideriamo la distribuzione della variabile casuale $W(t)$ al tempo t che significa

$$F_t(a) = Pr[W(t) \leq a] \tag{1.3}$$

Il processo stocastico (o processo aleatorio) è un vettore formato da un numero infinito di variabili casuali $\Pi = \{W_t | t \in T\}$ con $T \subset R$ definite su di uno spazio di probabiità (Ω, F, Pr). Qui T è un insieme enumerabile, cioè $t \in T$ ha valori assimilabili ad un insieme di numeri naturali equispaziati

1) $T = \{\cdots, -7, -6, -2, -1, 0, 2, 8, 9, 10, \cdots\} \cdots,$

$$W_{-7}, W_{-6}, W_{-2}, W_{-1}, W_0, W_2, W_8, W_9, W_{10}, \cdots$$

2) $T = \{\cdots, 3, 6, 9, 12, 15, \cdots\}$

$$\cdots, W_3, W_6, W_9, W_{12}, W_{15}, \cdots$$

3) $T = \{\cdots, -4, -3, -2, -1, 0, 1, 2, 3, 4, \cdots\} \cdots$

$$W_{-4}, W_{-3}, W_{-2}, W_{-1}, W_0, W_1, W_2, W_{3,4}, \cdots$$

La w_t è l'osservazione effettuata al tempo t dove t è l'istante nel quale si è manifestato o è stato regitrato il valore del processo.

La figura che segue è una porzione dell'elettrocardiogramma acquisito da uno dei molti canali simultaneamente attivi sul corpo del paziente. dal punto di vista delle serie storiche l'obiettivo è di predisporre un modello in grado di anticipare eventuali episodi di pericolo.

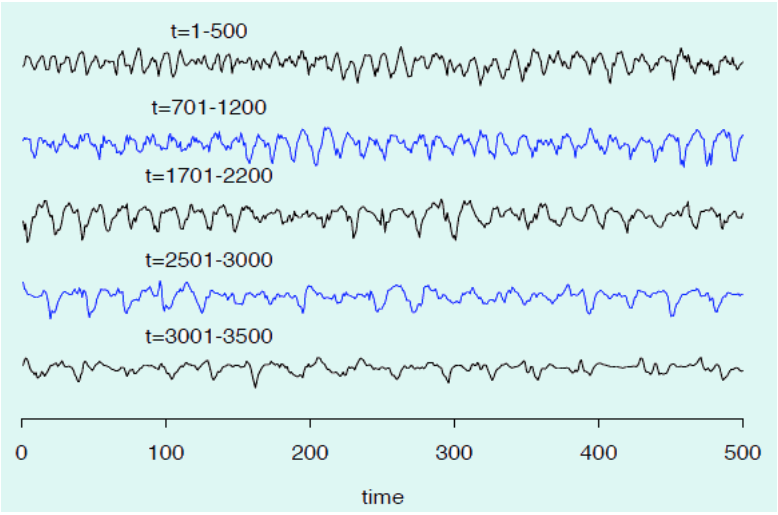


Figura 1.3: Grafico da un canale di ECG

La W_t è la variabile casuale che ha generato w_t, $F(W_t)$ è la funzione di ripartizione della singola variabile casuale con cui si esprimono le probabilità che W_t ricada in un certo intervallo ricadente nell'insieme di insiemi descritto dall'algebra Ω. In senso lato, si può dire che la serie storica è figlia del processo stocastico in quanto, per ciascuna delle infinite variabili casuali che formano il processo, la serie storica ne contiene una singola realizzazione.

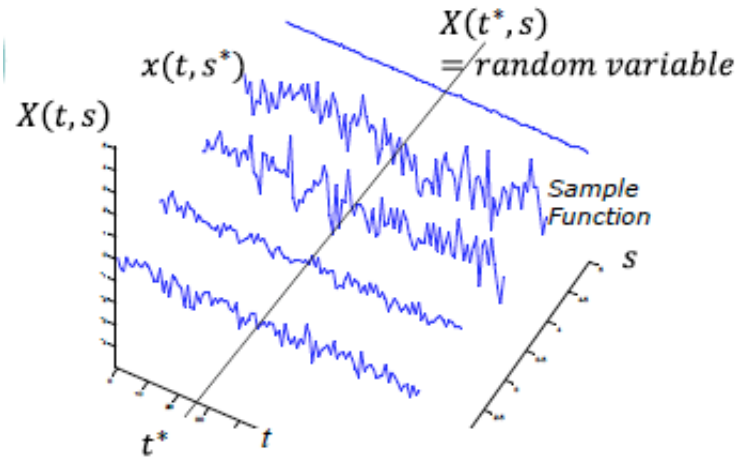


Figura 1.4: Processi stocastico e sue realizzazioni

Di solito, l'ordinamento delle variabili è regolare nel senso che gli scarti temporali rimangono costanti nel tempo e le W_t sono disposte ad intervalli equispaziati di tempo: $t_0, t_0+h, t_0+2h, \cdots, t_0+nh \rightarrow W_1, W_2, \cdots, W_n$. La metodologia studiata può essere estesa, con i dovuti accorgimenti, alle serie irregolari.

Come si è più volte affermato, ogni elemento del processo stocastico è una variabile casuale. E' utile, a questo punto, definire la funzione di ripartizione congiunta che coinvolge le infinite variabili casuali del processo

$$Pr\left[\cdots w_{t-1} \leq W_{t-1}, w_t \leq W_t, w_{t+1} \leq W_{t+1} \cdots\right] = F\left[\cdots, W_{t-1}, W_t, W_{t+1}\right] \quad (1.4)$$

Al momento sappiamo che le componenti del processo non sono indipendenti (altrimenti lo studio delle serie storiche avrebbe poco senso); infatti, il fatto che a_t raggiunga un valore elevato non induce alcuna reazione in a_{t+1} che quindi può tanto essere maggiore o minore rispetto ad a_t, il processo è privo di persistenza o isteresi e dunque non è utile a fini previsivi.

Un processo stocastico è noto quando si conoscono

- Le funzioni di ripartizione marginali cioè delle singole W_t;
- La funzione di ripartizione congiunta di tutte le possibili coppie (W_t, W_s)
- La funzione di ripartizione congiunta di tutte le possibili terne (W_t, W_s, W_r);
- $\cdots$ $\cdots$ $\cdots$

- La funzione di ripartizione congiunta di ogni n-tupla $(W_1, W_2, \cdots, W_n)$ per tutti gli n e per $n \to \infty$.

Tali conoscenze ben difficilmente possono realizzarsi ed in realtà il nostro interesse non è proprio per l'intera struttura probabilistica del processo, ma si limita invece ai possibili legami lineari tra le componenti aletorie in esso inserite. In particolare, qualora esistano finiti alcuni momenti delle variabili casuali, possiamo considerare

◦ Il livello medio al tempo t

$$E(W_t) = \int_{-\infty}^{+\infty} W_t dF_t = \mu_t \tag{1.5}$$

◦ La varianza al tempo t

$$Var(W_t) = E(W_t - \mu_t)^2 = \int_{-\infty}^{+\infty} 1(W_t - \mu_t)^2 dF_t = \sigma_t^2 \tag{1.6}$$

◦ La covarianza di ritardo (o *lag*) k

$$Cov(W_t, W_{t-k}) = E[(W_t - \mu_t)(W_{t-k} - \mu_{t-k})] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (W_t - \mu_t)(W_{t-k} - \mu_{t-k}) dF_t dF_{t-k} = \gamma(t, t-k) \tag{1.7}$$

in cui $t = 1, 2, \cdots, n, \cdots$ e $k = \pm 1, \pm 2, \cdots, \pm n, \cdots$ Il termine *lag* indica la differenza tra gli indici di posizione delle variabili coinvolte nel calcolo della covarianza.

1.1.3 La serie storica

La serie storica (o serie termporale o *time series*) è la realizzazione di un processo stocastico prodottasi considerando una osservazione da ciascuna variabile casuale del processo. La notazione $\{\cdots, w_1, w_2, \cdots, w_t, \cdots, w_n, \cdots\}$ indica una serie storica generata dal processo $\{\cdots, W_1, W_2, \cdots, W_t, \cdots, W_n, \cdots\}$. Come il processo stocastico, anche la serie storica, è formata da infiniti termini. Potremo ovviamente osservarne solo un numero finito. Un esempio è descritto nella figura 1.3.

La serie storica è la realizzazione osservata di un processo stocastico nel periodo da $t = 70$ a $t = 83$. La lunghezza della serie ovvero il numero dei suoi termini è $n = 83 - 70 + 1 = 14$. Il processo, nello stesso periodo, avrebbe potuto dare luogo a infinite altre realizzazioni e queste, a loro volta, ad infinite altre serie storiche.

Come la variabile casuale assegna una probabilità a ciascun risultato o intervalli di risultati del suo dominio, il processo stocastico assegna una funzione temporale campionaria (alias serie storica) $W(t, s)$

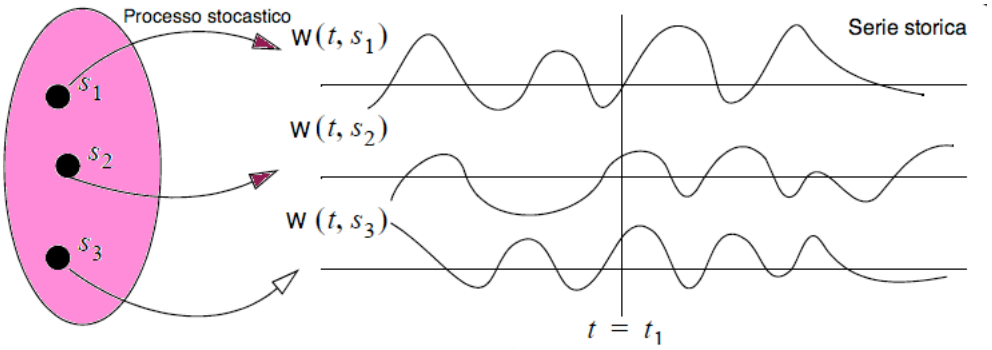


Figura 1.5: Molteplicità delle serie storiche originatesi in un processo

L'auspicio degli operatori è che la struttura di probabilità che governa il processo stocastico permetta di estendere a tutte queste eventuali realizzazioni, le conclusioni già raggiunte nello studio della singola serie storica osservata. In verità, le serie storiche, per costruzione, non sono mai osservabili nella loro interezza. Costituiscono però la base empirica per calcolare le previsioni o effettuare il monitoraggio sul processo stocastico. Questo è però possibile solo se la struttura di probabilità rimane stabile nel tempo, cioè non cambia o cambia molto lentamente o in maniera prevedibile da un periodo all'altro.

Nella figura 1.6 sono presentati sei diversi gruppi di serie storiche di cui alcuni molto simili tra di loro. In questi casi solo tecniche particolarmente sensibili possono ricostruire il corretto processo dalla serie storica osservata.

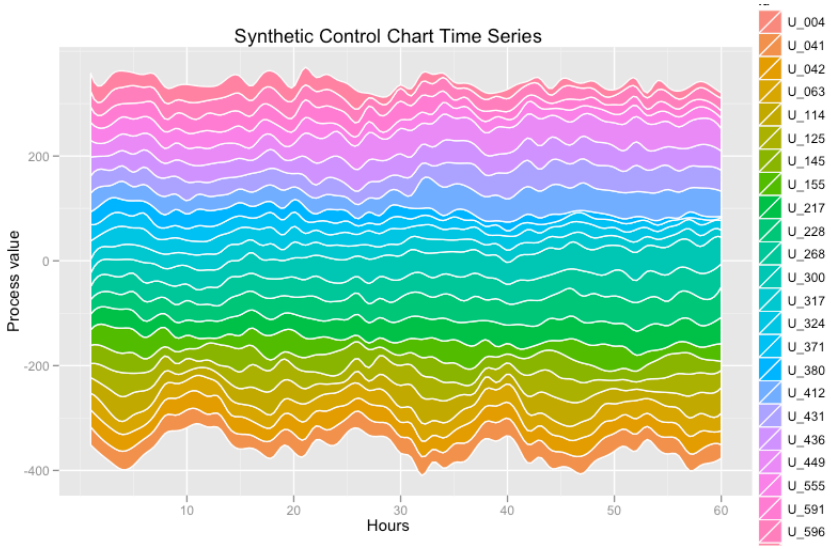


Figura 1.6: Clusters di serie storiche

E' bene sottolineare che ci occuperemo solo di serie storiche univariate cioè baseremo la conoscenza dei valori futuri di un fenomeno a ciò che è noto di una sola variabile. Inoltre, dovremo ignorare, per i limiti del programma, le interazioni tra serie temporale e articolazione territoriale delle osservazioni (modelli *space-time*).

E' una limitazione, certo, che contrasta con quello che sappiamo della realtà fatta da molteplici fenomeni che interagiscono e si influenzano a vicenda. D'altra parte apprendere le metodologie delle serie storiche multivariate richiede dei tempi di apprendimento lunghi e che, per le conoscenze e gli approcci della didattica ora conosciute, non può prescindere dalla competenza nell'approccio univariato. Questo però deve essere considerato solo una fase per arrivare a modelli più complessi.

1.1.4 Processo stocastico stazionario

Il processo stocastico è stazionario in senso forte (o, come anche si dice, in senso stretto) se la distribuzione di probabilità congiunta associata alla serie storica $\{W_{t_1}, W_{t_2}, \cdots, W_{t_n}\}$ è la stessa, qualunque sia la n-tupla degli indici $(t_1, t_2, \cdots, t_n)$.

$$F(W_{t_1}, \cdots, W_{t_n}) = F(W_{t_k+1}, \cdots, W_{t_k+n}) = F(W_{t_k-1}, \cdots, W_{t_k-n}) \tag{1.8}$$

ovvero è stazionario quel processo la cui funzione di ripartizione $F(.)$ rimane invariata quando ci si sposta indietro o in avanti sull'asse temporale rispetto al riferimento k. Ne consegue che, la stazionarietà in senso stretto implica delle variabili casuali identicamente distribuite e dunque con stessa media e stessa varianza (non necessariamente indipendenti se vogliamo proporre modelli che abbiano poi un senso pratico). E' ovvio che il fatto di essere uguali in media ed in varianza non ha alcuna influenza sulla loro esistenza finita o sull'esistenza finita dei momenti di ordine superiore.

La figura 1.7 è stata realizzata in base alla serie mensile della concentrazione media di $CO2$ rilevata all'osservatorio di Mauna Loa (Hawaii)

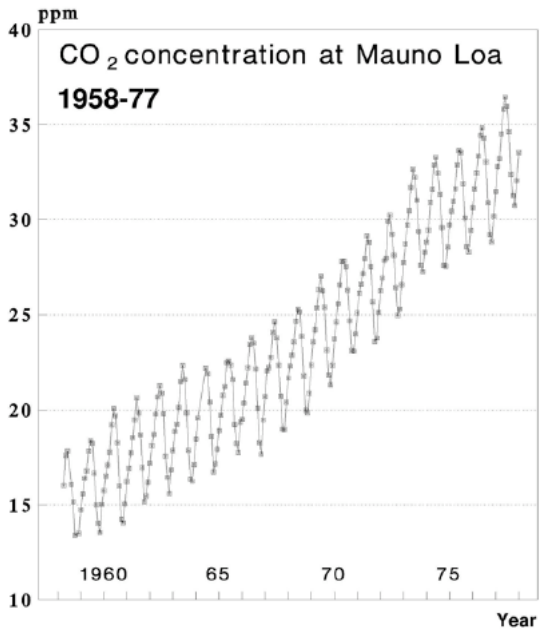


Figura 1.7: Serie storica non stazionaria

La serie storica si può inquadrare come la sovrapposizione di un processo stazionario ad un trend lineare ed una oscillazione periodica mensile. Il trend è sempre fonte di non stazionarietà, la periodicità stagionale non necessariamente risulta esplosiva.

Supponiamo che $n = 1$. In questo caso il processo stocastico è stazionario in senso forte se la funzione di ripartizione $F(W_t)$ rimane la stessa qualunque sia il tempo t cioè le distribuzioni marginali del processo, ottenute per i diversi valori di t, sono uguali. Quindi l'indice t può essere omesso dalla descrizione dei parametri del processo poiché risulterebbe irrilevante. Avremo, di conseguenza

$$E(W_t) = \int_{-\infty}^{+\infty} W_t dF_t = \mu$$

$$Var(W_t) = E(W_t - \mu_t)^2 = \int_{-\infty}^{+\infty} 1(W_t - \mu_t)^2 dF_t = \sigma^2 \quad (1.9)$$

Il processo stocastico stazionario in senso forte ha media costante ed è omoschedastico (varianze uguali indipendentemente dal tempo t, almeno condizionatemente a quello che è noto al tempo t. La costante μ rappresenta il livello intorno al quale il processo è in equilibrio ovvero il valore a cui tenderebbe se, per qualche ragione, ne venisse allontanato; infatti, μ è il valore - finito - che si configura allorché tutte le variabili diventano simultaneamente nulle. Inoltre

$$Var(W_t) = E(W_t - \mu)^2 = \int_{-\infty}^{+\infty} (W_t - \mu)^2 dF = \sigma^2 \quad (1.10)$$

La costante $\sigma_W^2 <$ inf esprime la variabilità del processo rispetto al suo livello di equilibrio. Da notare che oltre alla media ed alla varianza, anche gli altri momenti, finiti o non finiti, delle variabili casuali incluse nel processo stocastico stazionario in senso forte, non dipendono da t. E' importante ribadire che in una serie storica non stazionaria, valori molto distanti nel passato conservano una influenza che può essere pari o anche maggiore delle osservazioni più recenti.

L'ipotesi di stazionarietà implica, tra l'altro, il fatto che la funzione di ripartizione congiunta di due variabili casuali $F(W_t, W_{t+k})$ sia la stessa qualunque sia il ritardo k per cui si distanziano. Di queste distribuzioni doppie, l' aspetto che ci interessa più da vicino è la covarianza

$$Cov(W_t, W_{t-k}) = E[(W_t - \mu)(W_{t-k} - \mu)] =$$

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (W_t - \mu_t)(W_{t-k} - \mu_{t-k}) dF dF = \gamma(k) \quad (1.11)$$

che, quindi, nei processi stazionari, dipende solo da k e non da t. La $\gamma(k)$ è detta autocovarianza di *lag* k e rappresenta il valore atteso del co-movimento lineare della variabile casuale W_t rispetto alla $W_{t\pm k}$. Nella figura 1.8 sono riportare le autocovarianze fino al ritardo o *lag* 4

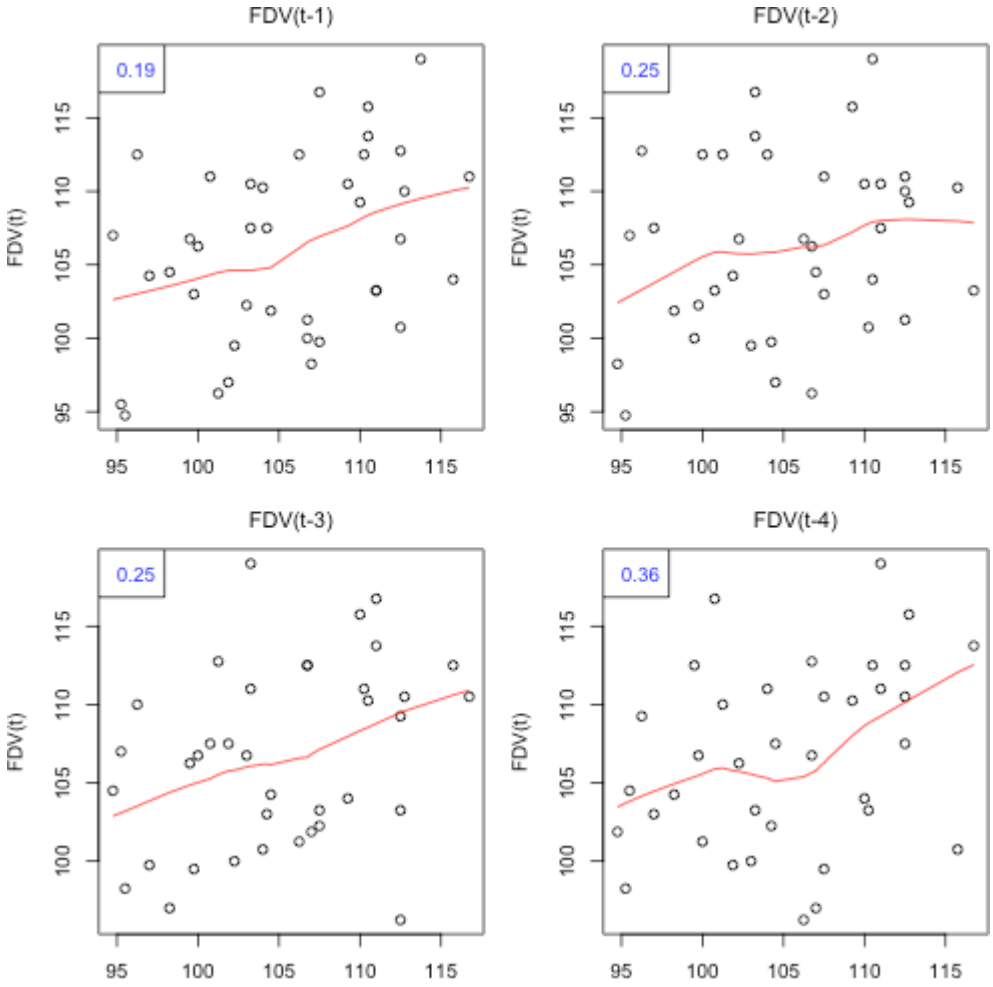


Figura 1.8: Andamento delle autocovarianze

Valori positivi di $\gamma(k)$ si ottengono se, tendenzialmente, valori superiori alla media del processo nella variabile in posizione t si accompagnano con valori superiori alla media della variabile in posizione $t+k$ ovvero $t-k$; lo stesso se si manifesta la tendenza delle due variabili ad avere entrambe valori inferiori alla media. I valori negativi di $\gamma(k)$ arrivano se le due variabili manifestano oscillazioni intorno alla media in modo cotraddittorio.

1.1.5 Il correlogramma

La autocovarianza dipende dall'unità di misura con cui è acquisita la variabile W_t. Per evitare di che l'analisi si complichi a causa di fattori di cui non è indipensabie prendersi cura quali la scala di misurazione del fenomeno, la serie storica si può standardizzare $Z_t = (W_t - \mu)/\sigma$ che hanno media forzata allo zero e varianza forzata all'unità. In questo caso, la autocovarianza coincide con la autocorrelazione

$$Cov(Z_t, Z_{t-k}) = E[(Z_t - \mu)(Z_{t-k} - \mu)] = \frac{\gamma(k)}{\sigma^2} = \rho(k) \tag{1.12}$$

che è indipendente da tempo t per i processi stazionari in senso stretto. La funzione di autocovarianza, e quindi di autocorrelazione, è una funzione pari, cioè $\gamma(k) = \gamma(-k)$ e $\rho(k) = \rho(-k)$; Inoltre, $\gamma(0) = \sigma^2$ per cui $\rho(0) = 1$. Ovviamente avremo anche $|\rho(k)| \leq 1$ dimostrabile grazie alla disuguaglianza di Cauchy-Schwartz.

La funzione di autocorrelazione è più pratica perchè essa riguarda il legame lineare tra variabili standardizzate depurate dalla unità di misura del fenomeno e con valori in un intervallo limitato: $[-1, 1]$ con significato univoco negli estremi. Infatti, $\rho(k) = \pm 1$ significa che tra le variabili di ritardo k esiste una relazione lineare esatta ascendente quando si raggiunge il limite superiore uno e discendente quando si raggiunge il limite inferiore -1. La rappresentazione grafica della funzione di $\rho(k)$ si realizza con il correlogramma: un grafico che collega le autocorrelazioni ai ritardi in cui si sono osservate:

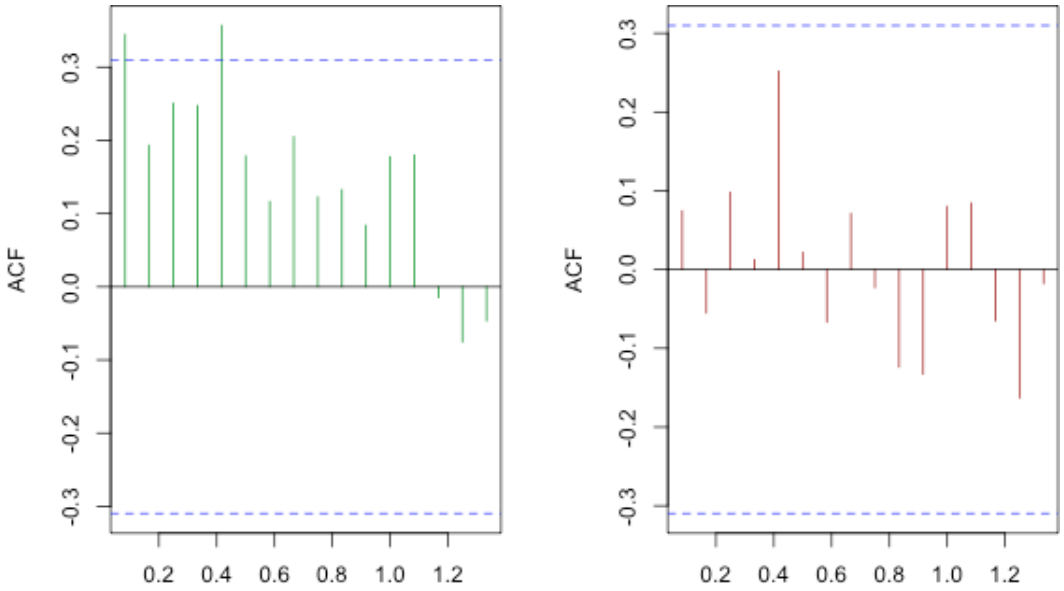


Figura 1.9: Esempi di correlogrammi

Conviene aggiungere un suggerimento: l'analisi delle autocorrelazioni si può intraprendere se la serie storica osservata contiene almeno una cinquantina di periodi. Se nel processo c'è isteresi o persistenza o inerzia, questa si rifletterà nella tendenza delle autocorrelazioni a manternesi elevate anche per ritardi elevati; se sono tutte prossime allo zero (tranne, ovviamente, quella a *lag* zero che è uno per definizione, si è di fronte ad una sequenza di valori ben poco informativa. La serie storica quindi trova la sua sintesi nel correlogramma in cui ciascuno elemento rappresenta la tendenza a variare in eccesso o in difetto di due variabili distanziate di un certo intervallo di tempi, rispetto alla media costante del processo. E' un riassunto univoco ed esaustivo? E' vero che, noto il correlogramma, diventa nota anche la serie storica ed il processo che l'ha generata? E' possibile risalire dal correlogramma alla serie storica e quindi al processo stocastico che l'ha generata?

Le autocovarianze e le autocorelazioni di una serie storica finita di n osservazioni

$$\{W_1, W_2, \cdots, W_t, \cdots, W_n\} \tag{1.13}$$

possono essere raccolte in una matrice $n \times n$:

$$\Gamma = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{n-2} & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{n-3} & \gamma_{n-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{n-4} & \gamma_{n-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \gamma_{n-1} & \gamma_{n-2} & \gamma_{n-3} & \cdots & \gamma_1 & \gamma_0 \end{bmatrix} \tag{1.14}$$

$$\mathbf{R} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-2} & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-3} & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{n-4} & \rho_{n-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & \rho_1 & 1 \end{bmatrix} \tag{1.15}$$

Queste matrici sono, per costruzione, semidefinite positive. Infatti, sia dato un vettore di costanti non tutte nulle $\boldsymbol{\lambda}$ e si definisca la combinazione lineare

$$c = \lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_{n+1} x_{n+1} = \boldsymbol{\lambda}^t \mathbf{x} \tag{1.16}$$

Allora la varianza di c si ottiene da $V(c) = \boldsymbol{\lambda}^t \Gamma \boldsymbol{\lambda}$ e poichè la varianza è sempre maggiore o uguale di zero, la matrice Γ deve necessariamente essere semidefinita positiva. Lo stesso vale per la matrice delle autocorrelazioni: $\boldsymbol{\lambda}^t \mathbf{R} \boldsymbol{\lambda}$.

Una matrice positiva semidefinita ha tutti i suoi minori principali di qualsiasi ordine maggiori o uguali a zero ovvero tutti i suoi autovalori sono non negativi. In particolare per $n = 2$ si ha

$$\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix} \tag{1.17}$$

cosicché $1 - \rho_1^2 \geq 0$ implica che $-1 \leq \rho_1 \leq 1$. Per $n = 3$ si devono verificare le condizioni seguenti:

$$\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix} \geq 0 \; ; \qquad \begin{vmatrix} 1 & \rho_2 \\ \rho_2 & 1 \end{vmatrix} \geq 0 \; ; \qquad \begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix} \geq 0 \tag{1.18}$$

che implicano $-1 \leq \rho_1 \leq 1$, $-1 \leq \rho_2 \leq 1$, $\rho_2 \geq 2\rho_1^2 - 1$. I valori delle due aucorrelazioni debbono ricadere nella regione indicata in verde nella figura 1.10

La conseguenza è che un insieme di valori in $[-1, 1]$ deve rispettare certi vincoli affinché possa essere considerato una funzione di autocorrelazione.

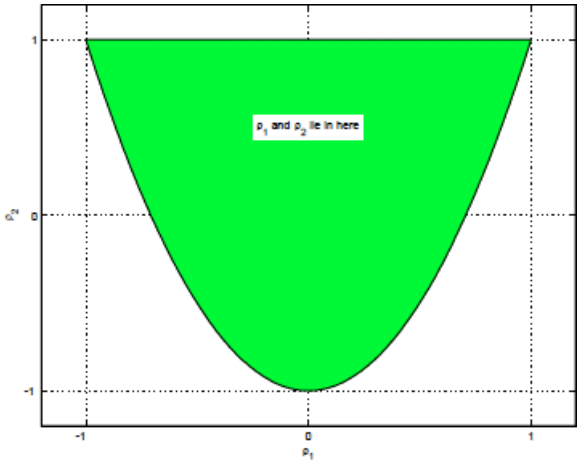


Figura 1.10: Regione di esistenza di una valida funzione di autocorrelazione per $n = 2$

In altre parole, non basta che le autocorrelazioni siano confinate all'intervallo tra meno uno ed uno, ma occorre che rispettino dei limiti più stringenti determinati. Ad esempio, per $n = 3$, se succede che $\rho_1 = 1/\sqrt{2}$, allora l'aurtocorrelazione di grado 2, cioè ρ_2 non può essere negativa all'interno di un correlogramma per una serie storica di lunghezza $n = 2$.

1.1.6 Processo stocastico stazionario in senso debole

La stazionarietà in senso forte è una condizione che, come si è detto, è ben difficile da realizzarsi od anche solo da verificare. Si preferisce quindi limitare l'analisi delle serie storiche a processi sottoposti a condizioni meno restrittive. Un processo stocastico è stazionario in senso debole o in senso ampio se il livello medio, la varianza e non dipendono dal tempo t e sono finite laddove, l'autocovarianza, dipende solo dal *gap* tra i ritardi. Se cioè

$$E(W_t) = \mu; \quad V(W_t)\,\sigma^2; \quad Cov(W_t, W_{t-k}) = \gamma(k) \tag{1.19}$$

Queste condizioni implicano che ciascuna delle variabili casuali inserite nel processo sia dotata dei momenti finiti fino al secondo ordine, ma non impongono altre restrizioni sulle funzioni di ripartizione congiunte o sui momenti di ordine superiore. La stazionarietà ha anche come conseguenza l'uguaglianza $\gamma(k) = \gamma(-k)$.

Poniamo $Y_t = \sum_{j=0}^{t} W_j$ e che le W_j siano incorrelate e parte di un processo con media nulla e varianza costante σ^2. Sia inoltre $Y_0 = 0$. Per ogni $k < 0$ la covarianza tra due termini della serie Y_t è data da

$$\gamma(k) = cov(Y_t, Y_{t+k}) = cov\left(\sum_{i=1}^{t} W_i, \sum_{i=1}^{t+k} W_j\right) \tag{1.20}$$

Poiché $cov\left(W_i, W_j\right) = 0$ se $i \neq j$, i termini delle somme coincideranno solo fino t e quindi si ottiene

$$\gamma\left(k\right) = cov\left(Y_t, Y_{t+k}\right) = var\left(\sum_{i=1}^{t} W_i\right) = \sum_{i=1}^{t} var\left(W_i\right) = \sum_{i=t}^{t} \sigma^2 = t\sigma^2 \tag{1.21}$$

In questo caso la autocovarianza di *lag* t dipende da t e dunque il processo non è stazionario.

Come altro esempio, ragioniamo sul processo $X_t, t = 1, 2, \cdots$, che, come nel caso precedente, ha media nulla e varianza costante σ^2. Se ora definiamo

$$W_t = \beta t + X_t \tag{1.22}$$

in cui abbiamo introdotto un *trend lineare*. L'aspettativa del processo è $E\left(W_t\right) = E\left(\beta t\right) + E\left(X_t\right) = \beta t + 0 = \beta t$ che non è stazionario poiché la media dipende dal tempo t.

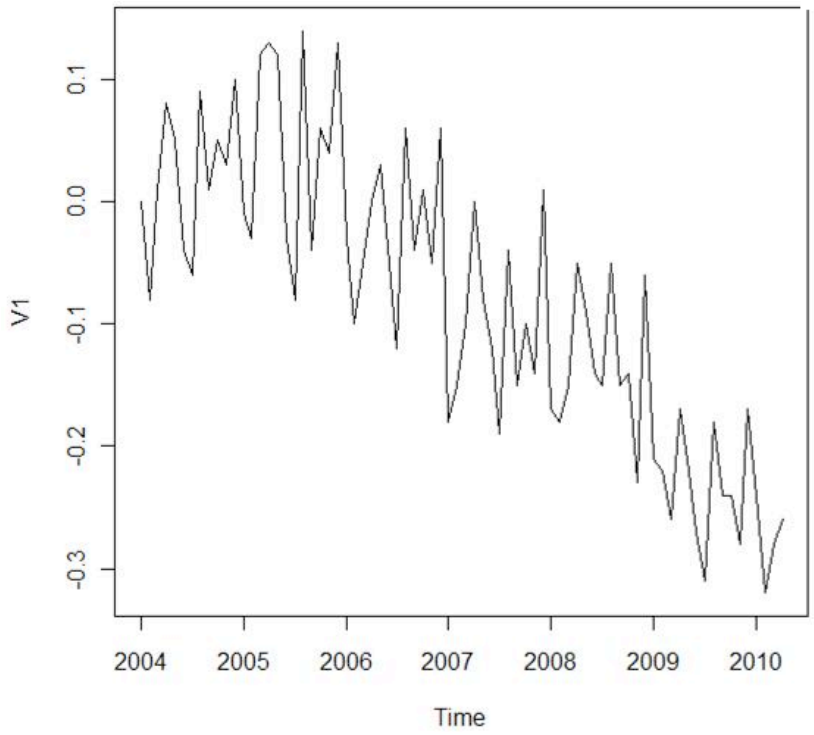


Figura 1.11: Serie storica non stazionaria in media a causa di un trend

In termini più semplici, la stazionarietà debole richiede che solo media e varianza calcolate in un qualsiasi sottoperiodo risultino uguali a quelle di un qualsiasi altro sottoperiodo. Ad esempio, la serie storica descritta nel grafico 1.12 non può essere considerata stazionaria in senso debole (tanto meno in senso stretto) in quanto i quattro segmenti in cui è stata divisa presentano differenze troppo nette. Il livello medio sembra modificarsi nei sottoperiodi considerati.

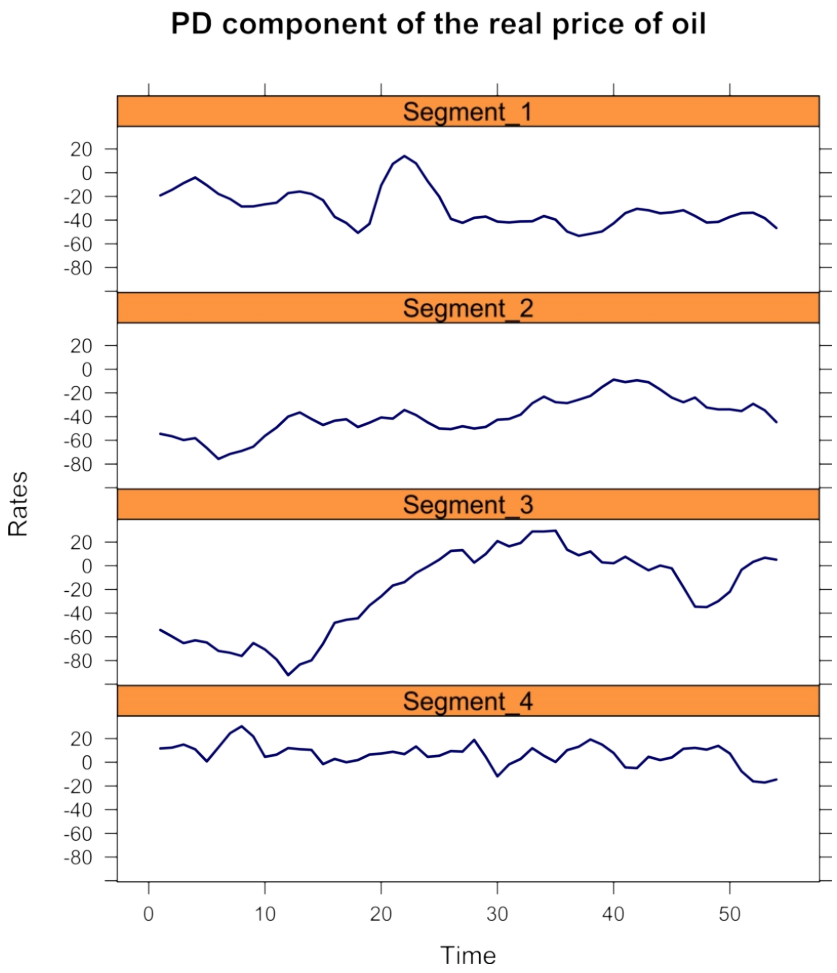


Figura 1.12: Precautionary demand component of the real price of oil

In generale, è molto difficile che le serie storiche osservate nella realtà manifestino stazionarietà e, pertanto, è necessario operare alcune trasformazioni che consentano di soddisfare i requisiti di omogeneità temporale (si veda il paragrafo 2.7 nel secondo capitolo). Una conseguenza immediata è che, se il processo è stazionario in senso debole, la conoscenza della sue funzione di auto-correlazione implica la conoscenza completa del processo ovvero, dato il processo, si può pervenire solo ad una specifica funzione di auto-correlazione. Inoltre, se è noto che dei processi sono stazionari in senso debole ed hanno la stessa funzione di auto-correlazione, allora i due processi coincidono.

La stazionarietà in senso forte non implica necessariamente la stazionarietà in senso debole. A questo fine basta notare che la stazionarietà in senso forte non presuppone una varianza finita: un processo formato da variabili casuali di Cauchy univariate, bivariate, etc. sarebbe stazionario in senso forte, ma non in senso debole dato che il valore atteso della Cauchy non esiste finito. Inoltre, un processo costruito con una trasformata non lineare di una variabile casuale che prima formava un processo strettamente stazionario, rimane strettamente stazionario. Invece, un pro-

cesso costruito con il quadrato di una variabile casuale che in scala naturale formava un processo stazionario in senso debole, non rimane necessariamente stazionario in senso debole. D'altra parte la stazionarietà in senso debole non può implicare quella in senso forte dato che l'assenza di legami con il tempo dei primi due momenti non costituisce garanzia per l'assenza degli stessi legami per i momenti di ordine più elevato.

Processi non stazionari e stazionari in senso debole o in senso forte possiedono caratteristiche e proprietà diverse a richiedono procedure inferenziali diverse. In prima approssimazione, si può ritenere che se una serie non mantiene un livello costante e/o una sequenza di oscillazioni di ampiezza regolare è molto probabilmente non stazionaria.

1.1.7 Processo *white noise*

La quasi totalità dei processi che incontreremo include variabili casuali correlate, quindi legate tra di loro almeno linearmente. Unica eccezione è il processo stazionario detto *white noise* o rumore bianco definito nel modo che segue

$$\begin{aligned} E(a_t) &= E(a_t|a_{t-1}, a_{t-2}, \ldots) = E(a_t|\text{ciò che è noto a t-1}) = 0 \quad (1.23) \\ V(a_t) &= V(a_t|a_{t-1}, a_{t-2}, \ldots) = V(a_t|\text{ciò che è noto a t-1}) = 0 \quad (1.24) \\ C(a_t, a_{t-k}) &= 0 \;\forall\; k \neq 0 \end{aligned}$$

Quindi i valori osservati in epoche diverse non sono correlati (per una stazionarietà in senso debole) o sono indipendenti (per una stazionarietà in senso forte), qualunque sia il ritardo che le differenzia. Ricordiamo che, in quanto stazionario, il processo *white noise* ha media finita e costante (di solito è nulla) e varianza altrettanto finita e costante. Senza tali condizioni non si potrebbe parlare di *white noise*. Da notare ancora che si parla di incorrelazione e non di indipendenza cioè si richiede di soddisfare un vincolo molto meno stringente.

La funzione di autocorrelazione è data da

$$\rho(k) = \begin{cases} 1 & \text{per } k = 0 \\ 0 & \text{altrimenti} \end{cases} \quad (1.25)$$

E' evidente che il processo è stazionario poiché nessuna delle componenti: media, varianza e covarianza variano con il tempo. Attenzione! perché è richiesto che la varianza sia finita e indipendente dal tempo non che essa sia costante. L'invarianza è necessaria per il valore atteso, ma processi eteroschedastici possono ancora essere stazionari.

Il processo *white noise* ha questo nome per l'analogia con la frequenza spettrale della *white light* i cui la densità spettrale della luce è distribuita sulla banda visibile in modo tale che l'occhio viene stimulato uniformemente dai tre colori fondamentali.

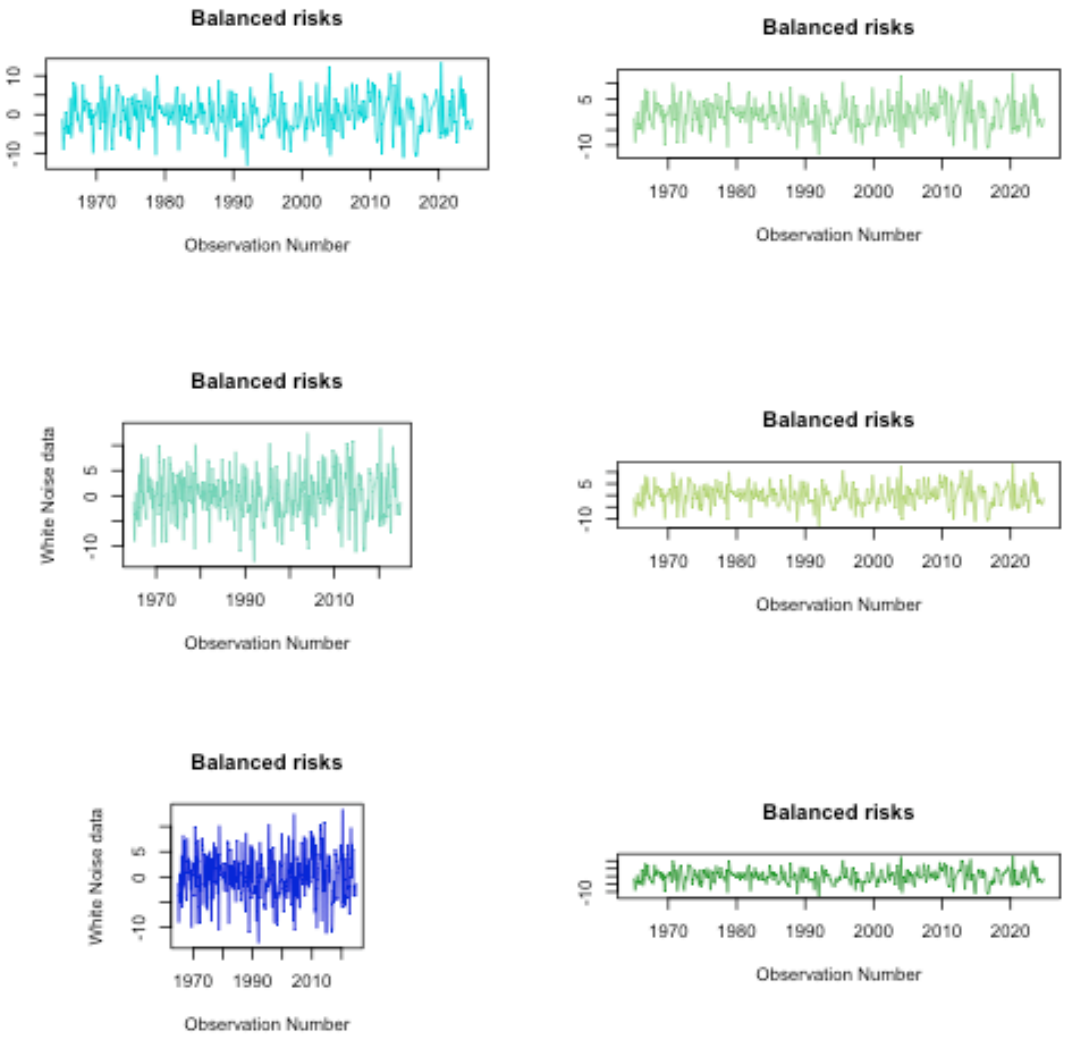


Figura 1.13: White noise

1.1.8 Processi gaussiani

Un processo stocastico si dice gaussiano quando la funzione di ripartizione congiunta di ogni sottoinsieme di variabili casuali in esso incluse, è multigaussiana. Come è noto, la distribuzione multigaussiana è determinata in modo univoco quando siano noti il vettore delle medie e la matrice di varianze-covarianze. Inoltre, laddove le variabili casuali del processo *white noise* generico sono solo incorelate, quelle del *white noise* gaussiano sono indipendenti (per il noto ruolo che hanno i coefficienti di correlazione nulli in una distribuzione congiunta multigaussiana). Per cui, dato che le medie e le varianze-covarianze sono finite e non dipendono da t, il processo gaussiano è stazionario in senso forte.

La diffusione dei processi gaussiani nella applicazioni pratiche dell'analisi delle

serie storiche si spiega con l'enorme beneficio computazionale e teorico che essi offrono al solo costo di una scarsa veridicità dell'ipotesi di distribuzione gaussiana delle variabili componenti il processo. Tutto si regge se le distribuzioni non si allontano troppo dal modello unimodale, simmetrico ed a code leggere. Se questo non succede, i vantaggi pratici e l'eleganza teorica non possono compensare per gli errori marchiani a cui si va incontro utilizzandoli nello studio di fenomeni sbagliati.

1.1.9 Ergodicità

La serie storica di ampiezza n è una realizzazione finita del processo stocastico. Le osservazioni che essa contiene non sono, almeno formalmente, un campione casuale di ampiezza n da una medesima variabile W, ma ogni osservazione è un campione di ampiezza 1 da ciascuna W_t. E' questo che consente la formazione di processi anche molto eterogenei con una distribuzione diversa ad ogni istante temporale. E' tuttavia possibile, sotto certe condizioni, considerare la serie storica come un campione di ampiezza n da un'unica variabile casuale. La condizione è quella di ergodicità.

Riprendiamo la legge dei grandi numeri di Kolmogorov. Essa stabilisce che se $X_t, t = 1, 2, \cdots, n$ sono delle variabili casuali indipendenti ed identicamente distribuite allora vale il seguente limite per la media globale n-esima delle variabili

$$\bar{X}_n = n^{-1} \sum_{t=1}^{n} X_t = \mu \tag{1.26}$$

Nell'ambito delle serie storiche, tuttavia, fronteggiamo medie di valori temporali non medie globali di valori della stessa natura. Per chiarire la differenza tra i due concetti consideriamo il seguente esperimento. Un gruppo di n particelle si muovono lungo una linea (l'asse reale); siamo interessati a seguire i loro spostamenti e di stabilire la posizione che ci si aspetta abbiano in un fissato istante di tempo. Se potessimo seguire ciascuna particella nelle sue evoluzioni, potremmo tracciarne i movimenti e stabilire la posizione media di ognuna di esse. Se immaginiamo che il gruppo di particelle si sia formato come un campione casuale di ampiezza n, allora le particelle potrebbero essere assimilate a delle variabili casuali i.i.d. e, a questo proposito, la legge dei grandi numeri assicura che la loro media (di tutte) converge alla aspettativa μ per $n \to \infty$. Tuttavia, come si è detto prima, nello studio delle serie storiche abbiamo una sola singola storia da narrare cioè siamo in grado di seguire una sola particella e solo di questa possiamo profilare l'andamento ed annotare la posizione raggiunta, diciamo $X_t, t = 1, 2, \cdots, T$. In particolare, la media dei valori nell'arco di tempo fino all'epoca T è $\bar{X}_T = T^{-1} \sum_{t=1}^{T} X_t$. Tale media (che in un contesto diverso definiremmo "media campionaria) non è la media di T realizzazioni di T variabili casuali distinte, ma la media di T osservazioni di una stessa variabile casuale X. Possiamo ritenere che la media così calcolata converga allo stesso limite come se si trattasse della media di tante particelle? La risposta è positiva solo se la serie storica è stazionaria ed ergodica.

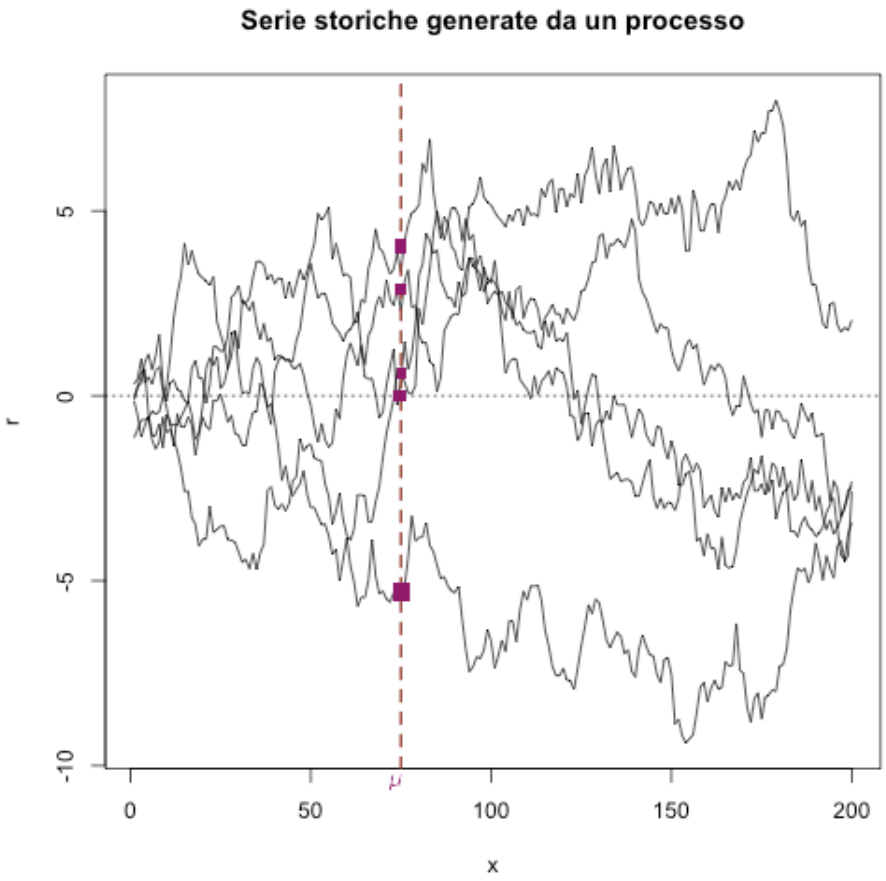


Figura 1.14: Esempio di ergodicità

Un processo stocastico si dice ergodico (rispetto ad una caratteristica $E\left[g\left(W_t\right)\right]$) quando la stima del parametro ottenuta considerando la serie storica un campione unico è consistente in media quadratica, converge cioè al valore vero del parametro da stimare secondo la legge.

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} W_t \xrightarrow{mq} E\left[g\left(W_t\right)\right] \tag{1.27}$$

Questo risultato richiede sia la stazionarietà che la ergodicità. La prima è già stata spiegata e si è visto che la stazionarietà non esclude la dipendenza all'interno della serie storica. Di converso, la ergodicità richiede **indipendenza asintotica tendenziale**. Ricordiamo che le definizioni date nei paragrafi precedenti sono tali che la stazionarità, di per sé, non garantisce la ergodicità.

La condizione di Hayashi che stabilisce la ergodicità di un processo già stazionario afferma che se f e g sono due funzioni limitate allora

$$\lim_{n\to\infty} \left|E\left[f\left(W_t,\ldots,W_{t+k}\right) g\left(W_{t+n},\ldots,W_{t+n+h}\right)\right]\right| = \left|E\left[f\left(W_t,\ldots,W_{t+k}\right)\right] E\left[g\left(W_{t+n}\ldots,W_{t+n+h}\right)\right]\right| \tag{1.28}$$

dove $f : R^k$ e $g : R^h$. In breve, questo significa che un processo stazionario è anche ergodico se le sue componenti sono anche asintoticamente, ma solo asintoticamente, indipendenti ovvero l'indipendenza aumenta all'aumentare del *gap* che separa due variabili casuali del processo. L'idea di fondo è quindi che la dipendenza nelle serie storiche non è del tipo presente/assente, ma graduale e non è sempre la stessa per tutti gli abbinamenti di variabili casuali. Man mano che le variabili casuali si distanziano all'intrno del processo, la dipendenza diminuisce.

Il fatto che si possa ottenere una stima consistente di una caratteristica del processo dallo studio di una sola realizzazione finita non è del tutto scontato. Nella figura 1.14 è illustrato il fatto che per un dato istante $t = i$ possiamo calcolare la media delle realizzazioni per le serie storiche originatesi nel processo

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} W_{t,i} \xrightarrow{mq} \mu \tag{1.29}$$

Se la media di insieme delle serie storiche tende alla media globale del processo, allora questo è ergodico rispetto alla media.

In poche parole, l'ergodicità implica che variabili allocate in posizioni molto discoste di uno stesso processo sono incorrelate (in questo senso il processo *white noise* è ergodico dato che le sue autocorrelazioni di *lag* diverso da zero sono tutte nulle).

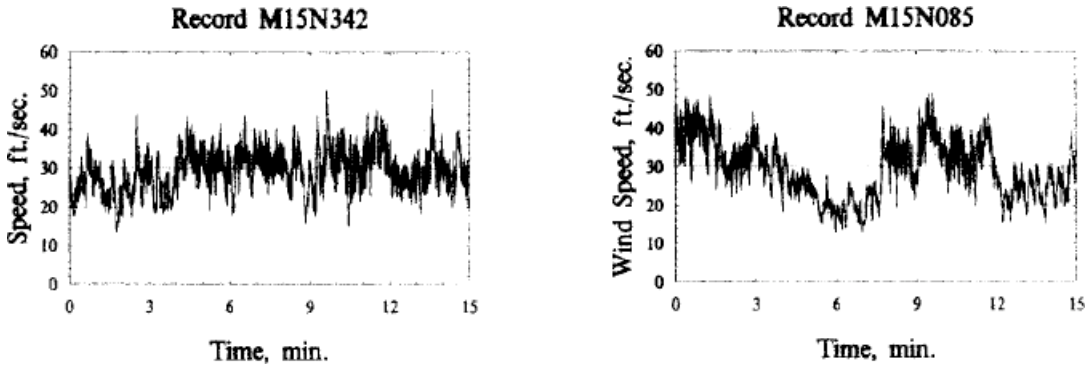


Figura 1.15: Wind records are classified as stationary (left) or nonstationary (right)

Esistono dei teoremi, che esulano comunque dall'ambito di questa introduzione alla metodologia Box-Jenkins, che permettono di stabilire la plausibilità dell'ipotesi di ergodicità per molti dei processi stazionari incontrati in pratica (in generale, comunque, l'ergodicità non implica la stazionarierà). Ad esempio, i processi stocastici gaussiani sono ergodici. Un altro esempio è la verifica delle radici unitarie che si riscontra nello studio dei processi stocastici lineari come base per una possibile cointegrazione di modelli per serie storiche diverse. Nel prosieguo studieremo solo processi ergodici rispetto ai momenti del primo e del secondo ordine anche se questo è più un atto di fede che una dimostrazione teorica perché sarà sempre molto difficile provare il contrario.

1.1.10 Ergodicità e autocorrelazioni

Un processo non è ergodico se la memoria degli eventi perdura per tanti periodi così da accumularsi al punto tale da impedire la decifrazione delle caratteristiche del processo utilizzando solo una sua realizzazione finita. Data la serie storica finita $\{W_1, W_2, \cdots, W_n\}$ proveniente da un processo stazionario ergodico, la media e la funzione di autocovarianza possono essere stimate in base alle relazioni

$$\begin{aligned} E(W_t) &= \hat{\mu} = \frac{\sum_{t=1}^{n} W_t}{n}; \qquad Var(W_t) = \hat{\sigma}^2 = \frac{\sum_{t=1}^{n} (W_t - \hat{\mu})^2}{n} \\ Cov(W_t, W_{t-k}) &= \hat{\gamma}(k) = \frac{\sum_{t=1}^{n} (W_t - \hat{\mu})(W_{t-k} - \hat{\mu})}{n} \end{aligned}$$

Nella formula, $\hat{\gamma}(k)$ è diviso per n invece che per $(n-k)$ per garantire la verifica dei vincoli stabiliti sulle autocovarianze teoriche. La stima della funzione di autocorrelazione si ottiene da

$$r_k = \frac{\hat{\gamma}(k)}{\hat{\sigma}^2}, \quad k = \pm 1, \pm 2, \ldots \tag{1.30}$$

Ad esempio per la serie storica $\{2, 4, 9, 5, 3, 9, 7, 2, 5, 4, 7, 3\}$ otteniamo

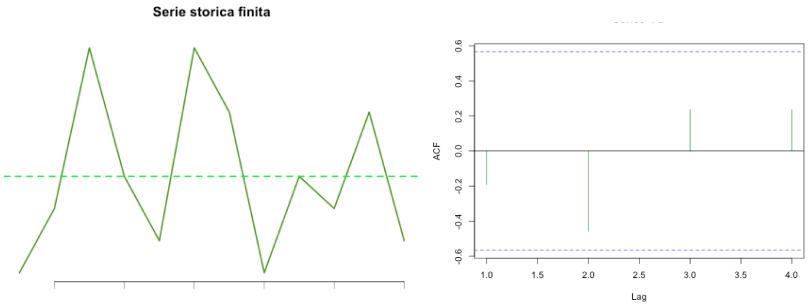


Figura 1.16: Serie storia breve e correlogramma

Per garantire sufficiente precisione alle stime, la lunghezza della serie storica finita dovrebbe essere piuttosto alta, diciamo superiore a 60. Inoltre, poichè il calcolo di $\hat{\gamma}(k)$ coinvolge sempre meno termini man mano che il ritardo aumenta, la qualità delle stime delle autocorrelazioni diminuisce all'aumentare del ritardo. Addirittura non potremo calcolare $\hat{\gamma}(k)$ per $k \geq (n-1)$; in genere, nel calcolo di $\hat{\gamma}(k)$ non si andrà oltre il lag $n/6$o altra soglia ritenuta congrua.

Per varie ragioni, ad esempio l'ergodicità, è opportuno verificare che i $\rho(k)$ tendano effettivamente a zero da un certo ritardo in poi. Poichè i $\rho(k)$ non sono noti si utilizzeranno a questo fine gli r_k osservati. Questi sono stimatori non distorti ovvero $E(r_k) = \rho(k)$ ed hanno una varianza che, nell'ipotesi di indipendenza, ha espressione

$$Var(r_k) \approx \frac{\sum_{i=-h}^{h} \rho(i)^2}{n} = \frac{1 + 2\sum_{i=1}^{h} \rho(i)^2}{n} = \sigma_k^2 \quad \text{per } k > h \tag{1.31}$$

Inoltre, se n è abbastanza grande, lo stimatore r_k ha distribuzione gaussiana con media $\rho(k)$ e varianza σ_k^2. Poiché anche σ_k^2 è incognita, verrà stimata sotituendo le autocorrelazioni osservate a quelle teroriche

$$s_k^2 = \frac{1 + 2\sum_{i=1}^{h} r_k^2}{n} \quad \text{per } k > h \tag{1.32}$$

Pertanto, la verifica di ipotesi

$$\begin{cases} H_0: & \rho(k) = 0 \\ H_1: & \rho(k) \neq 0 \end{cases} \tag{1.33}$$

potrà essere basata sulla gaussiana $N(0, s_k^2)$ sebbene questo sia formalmente corretto solo quando n è elevato. Quando H_0 è vera, i valori di $\rho(k)$ esterni all'intervallo $[-2 * s_k; 2 * s_k]$ possono verificarsi solo nel 5% dei casi; quindi se r_k non ricade in questo intervallo rifiuteremo l'ipotesi che il coefficiente di correlazione teorico $\rho(k)$ sia nullo e riterremo significativa la relazione lineare tra W_t e W_{t-k}. Una ulteriore approssimazione di σ_k^2 è costituita dalla $1/n$ che fornisce dei limiti di significatività: $[-2/\sqrt{n}, 2/\sqrt{n}]$ uguali per tutti i ritardi.

Poiché la serie storica ha un numero contenuto di osservazioni, l'attendibilità del calcolo delle autocorrelazioni diminuisce all'aumentare del ritardo considerato. Quante calcolarne? Ecco due regole molto empiriche per stabilire il *lag* massimo Q sono le seguenti:

$$\begin{aligned} Q &= \lfloor \sqrt{n} + 0.5 \rfloor + s \\ Q &= \lfloor (s * (n/100)^{0.25} + 0.5 \rfloor \end{aligned} \tag{1.34}$$

dove s è la periodicità stagionale. Ad esempio $s = 12$ per serie mensili e $s = 4$ per serie trimestrali.

1.1.11 Autocorrelazioni parziali

L'autocorrelazione semplice misura il legame lineare tra variabili collocate ad intervalli equispaziati del processo stocastico. Per un fissato valore d k si ha

$$\rho(k) = \frac{Cov(W_t, W_{t-k})}{\sigma^2}. \tag{1.35}$$

In particolare ρ_1 esprime il grado di dipendenza lineare tra variabili immediatamente contigue del processo. Dalla teoria della regressione lineare semplice sappiamo che ρ_1 coincide con il coefficiente angolare della retta ai minimi quadrati che lega W_t a W_{t-1} in caso di variabili standardizzate. Consideriamo il modello lineare semplice nel tempo t con errori *white noise*

$$(W_t - \mu) = \phi_1 (W_{t-1} - \mu) + a_t \tag{1.36}$$

Moltiplichiamo entrambi i lati della relazione (1.36) per $(W_{t-1} - \mu)$ per ottenere

$$(W_t - \mu)(W_{t-1} - \mu) = \phi_1 (W_{t-1} - \mu)^2 + (W_{t-1} - \mu) a_t \tag{1.37}$$

Determiniamo il valore atteso di entrambi i membri della (1.37)

$$\begin{aligned} E[(W_t - \mu)(W_{t-1} - \mu)] &= \phi_1 E\left[(W_{t-1} - \mu)^2\right] + E[(W_{t-1} - \mu) a_t] \\ \gamma_1 &= \phi_1 \sigma^2 \end{aligned} \tag{1.38}$$

dato che $E[(W_{t-1} - \mu) a_t] = 0$ per ogni $k \neq 0$ e la variabile W_{t-k} non coinvolge che termini a_t con ritardo diverso da t e questi, per definizione, sono incorrelati con a_t. La seconda relazione in (1.38) implica $\phi_1 = \gamma_1/\sigma^2 = \rho_1$.

Supponiamo ora di esprimere $(W_t - \mu)$ in funzione delle due variabili immediatamente precedenti

$$(W_t - \mu) = \phi_{12} (W_{t-1} - \mu) + \phi_{22} (W_{t-2} - \mu) + a_t \tag{1.39}$$

dove il secondo indice nei coefficienti indica il numero di variabili passate da considerare come esplicative di W_t. Ripetendo le stesse operazioni di prima sia con $(W_{t-1} - \mu)$ che con $(W_{t-2} - \mu)$ si arriva a

$$\begin{aligned} E[(W_t - \mu)(W_{t-1} - \mu)] = \phi_{12} E\left[(W_{t-1} - \mu)^2\right] + \\ \phi_{22} E[(W_{t-1} - \mu)(W_{t-2} - \mu)] + E[(W_{t-1} - \mu) a_t] \end{aligned} \tag{1.40}$$

$$\begin{aligned} E[(W_t - \mu)(W_{t-2} - \mu)] = \phi_{12} E[(W_{t-1} - \mu)(W_{t-2} - \mu)] + \\ \phi_{22} E\left[(W_{t-2} - \mu)^2\right] + E[(W_{t-2} - \mu) a_t] \end{aligned} \tag{1.41}$$

ovvero

$$\begin{aligned} \gamma_1 &= \phi_{12}\sigma^2 + \phi_{22}\gamma_1 \rightarrow \rho_1 = \phi_{12} + \phi_{22}\rho_1 \\ \gamma_2 &= \phi_{12}\gamma_1 + \phi_{22}\sigma^2 \rightarrow \rho_2 = \phi_{12}\rho_1 + \phi_{22} \end{aligned} \tag{1.42}$$

In generale, se si esprime $(W_t - \mu)$ in funzione delle k variabili casuali ad essa contigue nella scansione temporale, valgono le cosiddette equazioni di Yule-Walker efficacemente descritte da un sistema lineare.

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{k-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{1k} \\ \phi_{2k} \\ \phi_{3k} \\ \cdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \cdots \\ \rho_k \end{bmatrix} \tag{1.43}$$

La soluzione del sistema $\boldsymbol{\phi} = \mathbf{R}\boldsymbol{\rho}$ esiste quando la matrice delle autocorrelazioni è definita positiva (tutti i minori principali di qualsiasi ordine sono strettamente

positivi ovvero tutti gli autovalori sono maggiori di zero). In verità, del vettore delle soluzioni $\boldsymbol{\phi}$ a noi interessa solo l'ultimo elemento ϕ_{kk} perché indicativo del legame lineare tra la variabile in posizione t, cioè W_t e la variabile W_{t-k} in posizione $t-k$. Per $k=1$ avremo ovviamente $\phi_{11}=\rho_1$. Per $k=2$ si ha

$$\phi_{22}=\frac{\rho_2-\rho_1^2}{1-\rho_1^2} \tag{1.44}$$

In generale, per determinare ϕ_{kk} si può utlizzare la formula

$$\phi_{kk}=\frac{\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{k-4} & \rho_3 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{bmatrix}}{\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{k-4} & \rho_{k-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{bmatrix}} \tag{1.45}$$

I parametri $\phi_{kk}, k=1,2,\cdots$ sono detti coefficienti di autocorrelazione parziale e hanno un significato simile ai coefficienti del modello di regressione lineare multipla: il fatto che ϕ_{kk} sia significativamente diverso da zero implica che il termine W_{t-k} è utile per la "spiegazione" di W_t. Il legame che si instaura tra il ritardo k-esimo, e la correlazione parziale ϕ_{kk} costituisce la funzione di autocorrelazione parziale. E' da sottonolineare che le autocorrelazioni parziali sono in fondo delle trasformate lineari delle autocorrelazioni semplici e non aggiungono quindi nessuna altra informazione. Esse però agevolano, come si vedrà, il collegamento tra il correlogramma ed il processo stocastico che è all'origine della serie storica studiata.

La stima delle autocorrelazioni parziali può essere effettuata sfruttando pienamente l'analogia con il modello di regressione lineare e cioè considerando le stime ai minimi quadrati nel modello

$$(W_t-\mu)=\phi_{1k}\,(W_{t-1}-\mu)+\phi_{2k}\,(W_{t-2}-\mu)+\cdots+\phi_{kk}\,(W_{t-k}-\mu)+a_t \tag{1.46}$$

Per vari valori di k ed analizzando poi il coefficiente associato all'ultimo termine. In alternativa si possono usare le equazioni di Yule-Walker applicate alle autocorrelazioni osservate $\hat{\boldsymbol{\phi}}=\mathbf{R}^{-1}\mathbf{r}$.

Anche gli stimatori $\hat{\boldsymbol{\phi}}$ sono funzioni dei valori empirici e quindi soggetti ad errori. La loro varianza può essere approssimata con

$$Var\left(\hat{\boldsymbol{\phi}}\right) \approx \frac{1}{n} \tag{1.47}$$

quindi, nel grafico delle autocorrelazioni parziali potremo porre delle bande di confidenza al 95% con le quantità $\pm 2/\sqrt{n}$. Nella figura 1.17 sono riportate le autocorrelazioni parziali isieme agli intervalli di confidenza specifici per ogni ritardo, possibile grazie al miglioramento del software.

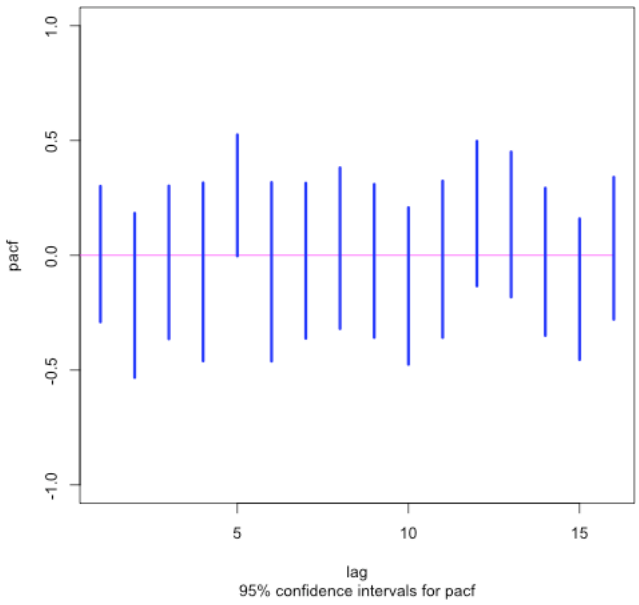


Figura 1.17: Esempi di correlazioni parziali

Se il segmento verticale tocca la linea orizzontale, l'autocorrelazione può essere considerata non significativa: meno centrato è il segmento, maggiore è la significatività.

CAPITOLO 2

Analisi preliminari

La stazionarietà discussa nel capitolo precedente implica una media costante ed oscillazioni intorno ad essa contenute e regolari e che non modificano tale comportamento con il passare del tempo. Come si è detto, sono ipotesi molto forti e raramente i fenomeni soggetti a variazioni temporali stocastiche riescono a rispettarle per intero (e, d'altra parte, non vi sono tenute né hanno uno speciale interesse ad obbedirvi). Ingenuamente si potrebbe essere convinti che basti cercare i dati per trovarli. In realtà, quando ci si trova di fronte ad indagini effettive e problemi concreti ci si accorge subito che la completezza dei dati è una mera illusione. C'è bisogno di "curare" la serie storica per renderla operativa, perché, se non si intervenisse, questo potrebbe rappresentare un *minus* per chi dovrà poi utilizzarla.

Chiunque abbia gestito osservazioni rilevate "in vivo si è trovato di fronte ai seguenti problemi:

1. Definizione coerente di unità, variabili e circostanze di rilevazione.
2. Valori anomali.
3. Valori mancanti.
4. Trasformazioni delle variabili.

Chi afferma di non avere avuto questi problemi ha solo evitato di dare un'occhiata più approfondita ai propri dati. Ad esempio, la scelta dell'unità di tempo può risultare cruciale ai fini del successo dell'analisi di una serie storica.

Installazione di linee telematiche interattive con le amministrazioni pubbliche. Dati giornalieri (1° grafico) e dati settimanali (2° grafico). L'andamento stagionale è più evidente nel profilo giornaliero; il trend è più nitido in quello settimanale.

Lo scopo del trattamento preliminare della serie storica è di rendere comparabili nel tempo le "unità di osservazione di modo che le differenze riscontrate siano univocamente identificabili con cambiamenti nel fenomeno: una variazione di temperatura deve significare un calo o un aumento della febbre, non un cambiamento del termometro. Attenzione! L'esigenza di intervenire sui dati per depurarli, pesar-

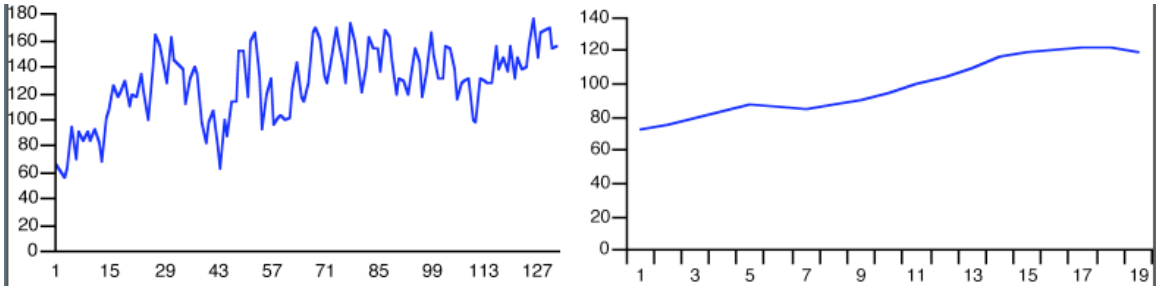


Figura 2.1: Cambio dell'unità di tempo

li, integrarli e trasformarli in indicatori usabili in economia non è sempre condivisa. Per alcuni, i dati dovrebbero rimanere intonsi e narrare per intero la loro storia: sintomi che il buon dottore saprà interpretare. Gli istituti centrali e i produttori di dati in genere, debbono invece preoccuparsi di fornire serie storiche massimamente complete ed omogenee che consentano agli utenti di concentrarsi sul loro utilizzo più che sulla loro costruzione.

In questo capitolo studieremo alcuni accorgimenti che possono spingere la serie osservata ad trasformarsi in un processo stazionario in media e varianza. Nella nostra impostazione quindi la cura preliminare della serie storica deve servire a renderla operativa per la costruzione dei modelli Box-Jenkins (che per noi è un fine). La finalità del fornitore di serie storiche è di proporre un vettore di dati attendibile e questo può spesso ottenersi applicando la metodologia Box-Jenkins (che qui è un mezzo).

2.1 Correzioni di calendario

La correzione preliminare per le componenti di calendario (giorni lavorativi lunghezza del mese o del trimestre) può essere effettuata in automatico. Nel caso di serie mensili occorre tenere conto della diversa lunghezza dei mesi.

$$Dato\ gennaio = \frac{Totale\ gennaio}{31}; \quad Dato\ febbraio = \frac{Totale\ febbraio}{28\ o\ 29} \tag{2.1}$$

Allo stesso modo, per i dati trimestrali si potrebbero considerare

$$Dato\ 1°\ Trim. = \frac{Totale\ 1°\ Trim.}{90\ o\ 91}; \quad Dato\ 2°\ Trim. \frac{Totale\ 2°\ Trim.}{91} \tag{2.2}$$

In alternativa, si può moltiplicare il dato per un fattore di correzione ricavato dalla formula:

$$correzione = \frac{Numero\ medio\ di\ giorni}{Numero\ di\ giorni} \tag{2.3}$$

Di volta in volta sarà l'unità di tempo della serie storica e la sua natura a suggerire le opportune correzioni di proporzionalità. Vale la pena sottolineare che il metodo

proporzionale, basato su un mero rapporto meccanico aprioristicamente valido sempre e ovunque, non dà risultati sempre soddisfacenti. Infatti, l'istituto di statistica italiano, dal dicembre 2002, lo ha accantonato.

L'utilità della procedura si avverte in questo passaggio pubblicato sul "Sole-24 ore: *Tuttavia, sempre secondo i dati diffusi questa mattina dall'Ufficio statistico federale, il Pil del secondo trimestre ha anche segnato, dopo le correzioni di calendario necessarie, un lieve incremento dello* $0,6\%$ *rispetto allo stesso periodo del 2004.*

Peraltro, il Regolamento del Consiglio dell'Unione Europea n. 1165/98 dispone che gli Stati membri trasmettano ad Eurostat i dati relativi alla produzione industriale corretti per i giorni lavorativi.

2.1.1 Effetto giorni lavorativi

Un anno tropico (tempo che intercorre tra due equinozi o solstizi dello stesso tipo) dura 365 giorni 6 ore e 3.6 minuti. Per tenere conto delle frazioni si aggiunge: un giorno ogni 4 anni (febbraio è di 29 giorni se l'anno è un numero divisibile per 4 (il 2008 è stato un anno bisestile perché $2008/4 = 502$ senza resto). Questa regola non vale se l'anno è divisibile anche per 100. Ad esempio il 1700 non è stato un'anno bisestile anche se era divisibile per 4. Il febbraio di 29 torna per gli anni che siano divisibili per $400 = 4 \times 100$ per tenere conto dei 3.6 minuti (infatti $3.6 \times 400/60 = 24$). Il 2000 è stato infatti un anno bisestile.

Se il primo gennaio cade di lunedì nell'anno non bisestile, il seguente cadrà di martedì. Se il primo gennaio cade di lunedì nell'anno bisestile, il seguente cadrà di mercoledì. La medesima struttura di calendario si ripete ogni $4 \times 7 = 28$ anni. L'effetto 'giorni lavorativi" è dovuto al numero di occorrenza dei giorni della settimana che cambia da un anno all'altro: numero di non festivi mensili per le importazioni, numero di sabati al mese per una città mercato.

L'effetto giorni lavorativi (EGL) è sistematico ed ha carattere di stabilità attraverso gli anni. E' possibile determinarlo dai dati mensili. La composizione dei mesi varia da un anno all'altro. Consideriamo 22 tipologie di struttura dei mesi: 7 tipologie di mesi a 31 giorni, 7 tipologie di mesi a 30 giorni e 8 tipologie di febbraio. Se si considerano le possibile coppie di mesi sono possibili 7 abbinamenti ciascuno per $31-31$, $30-31$, $31-30$, $31-28$, $28-31$, $31-29$, $29-31$ che possono verificarsi con una frequenza prefissata.

Queste strutture si ripetono ogni 28 anni. Ma esiste davvero l'effetto giorni lavorativi? Consideriamo un impianto che opera in per cinque giorni feriali settimanali e valutiamo due ipotesi

A) Un gennaio con 23 GL (4 sabati e 4 domeniche)

B) Un gennaio con 21 GL (5 sabati e 5 domeniche)

Quindi la differenza relativo in termini di giorni lavorativi è $(23-21)\,100/21 = 9.5\%$. Quindi, a parità di ogni altra condizione nel caso B) risulterebbe un calo della produzione esclusivamente dovuto agli effetti di calendario.

Tipo di mese	Primo giorno del mese	Tipo di giorno						
		Lunedì	Martedì	Mercoledì	Giovedì	Venerdì	Sabato	Domenica
					Mesi di 31 giorni			
1	Lunedì	5	5	5	4	4	4	4
2	Martedì	4	5	5	5	4	4	4
3	Mercoledì	4	4	5	5	5	4	4
4	Giovedì	4	4	4	5	5	5	4
5	Venerdì	4	4	4	4	5	5	5
6	Sabato	5	4	4	4	4	5	5
7	Domenica	5	5	4	4	4	4	5
					30 giorni (novembre-aprile-giugno-settembre)			
8	Lunedì	5	5	4	4	4	4	4
9	Martedì	4	5	5	4	4	4	4
10	Mercoledì	4	4	5	5	4	4	4
11	Giovedì	4	4	4	5	5	5	4
12	Venerdì	4	4	4	4	5	5	4
13	Sabato	4	4	4	4	4	4	5
14	Domenica	5	4	4	4	4	4	5
					Febbraio bisestile			
15	Lunedì	5	4	4	4	4	4	4
16	Martedì	4	5	4	4	4	4	4
17	Mercoledì	4	4	5	4	4	4	4
18	Giovedì	4	4	4	5	4	4	4
19	Venerdì	4	4	4	4	5	5	4
20	Sabato	4	4	4	4	4	4	4
21	Domenica	4	4	4	4	4	4	5
					Febbraio NON bisestile			
22	Qualsiasi	4	4	4	4	4	4	4

Figura 2.2: Tipologie delle coppie di mesi e giornate lavorative

Anno	Gennaio	Febbraio	Marzo	Aprile	Maggio	Giugno	Luglio	Agosto	Settembre	Ottobre	Novembre	Dicembre	Anno
2000	6	16	3	13	1	11	6	2	12	7	10	5	**1972**
2001	1	22	4	14	2	12	7	3	13	1	11	6	**1973**
2002	2	22	5	8	3	13	1	4	14	2	12	7	**1974**
2003	3	22	6	9	4	14	2	5	8	3	13	1	**1975**
2004	4	21	1	11	6	9	4	7	10	5	8	3	**1976**
2005	6	22	2	12	7	10	5	1	11	6	9	4	**1977**
2006	7	22	3	13	1	11	6	2	12	7	10	5	**1978**
2007	1	22	4	14	2	12	7	3	13	1	11	6	**1979**
2008	2	19	6	9	4	14	2	5	8	3	13	1	**1980**
2009	4	22	7	10	5	8	3	6	9	4	14	2	**1981**
2010	5	22	1	11	6	9	4	7	10	5	8	3	**1982**
2011	6	22	2	12	7	10	5	1	11	6	9	4	**1983**
2012	7	17	4	14	2	12	7	3	13	1	11	6	**1984**
2013	2	22	5	8	3	13	1	4	14	2	12	7	**1985**
2014	3	22	6	9	4	14	2	5	8	3	13	1	**1986**
2015	4	22	7	10	5	8	3	6	9	4	14	2	**1987**
2016	5	15	2	12	7	10	5	1	11	6	9	4	**1988**
2017	7	22	3	13	1	11	6	2	12	7	10	5	**1989**
2018	1	22	4	14	2	12	7	3	13	1	11	6	**1990**
2019	2	22	5	8	3	13	1	4	14	2	12	7	**1991**
2020	3	20	7	10	5	8	3	6	9	4	14	2	**1992**
2021	5	22	1	11	6	9	4	7	10	5	8	3	**1993**
2022	6	22	2	12	7	10	5	1	11	6	9	4	**1994**
2023	7	22	3	13	1	11	6	2	12	7	10	5	**1995**
2024	1	18	5	8	3	13	1	4	14	2	12	7	**1996**
2025	3	22	6	9	4	14	2	5	8	3	13	1	**1997**
2026	4	22	7	10	5	8	3	6	9	4	14	2	**1998**
2027	5	22	1	11	6	9	4	7	10	5	8	3	**1999**

Figura 2.3: Giornate lavorative negli anni

Un effetto di cui per il momento non teniamo conto (ma che srebbe utile approfondire) è la correzione della serie storica per tenere dell'effetto dovuto alle festività infrasettimanali.

2.1.2 Feste mobili (Effetto pasqua)

L'effetto di festività che cadono all'interno della settimana può esplicarsi in un aumento delle spese in attività turistiche o in una riduzione nella produzione. Non è chiaro come tenerne conto nella definizione del dato mensile. Spesso, le festività sono strettamente connesse alla composizione di calendario. Ad esempio, se Natale cade di lunedì, dicembre inizia di venerdì e contiene 5 sabati e 5 domeniche. La presenza di due festività in un mese non necessariamente ha l'effetto doppio di una festività sola. Questi effetti tendono in genere ad essere ignorati.

La festa mobile più importante nell'occidente cristiano è la Pasqua che cade in un giorno variabile tra il 22 marzo ed il 25 di aprile. Più la Pasqua è bassa e maggiore è il suo impatto sul mese di marzo. Una pasqua alta influirà essenzialmente sul mese di aprile. L'effetto nelle serie storiche può essere

a) IMMEDIATO: chiusura di alcune attività ed apertura di altre nel periodo pasquale.
b) GRADUALE: si manifeta e si estende anche nel periodo che precede ed in quello che segue.

anno	Pasqua	anno	Pasqua	anno	Pasqua
1970	29 marzo	1989	26 marzo	2008	23 marzo
1971	11 aprile	1990	15 aprile	2009	12 aprile
1972	2 aprile	1991	31 marzo	2010	4 aprile
1973	22 aprile	1992	19 aprile	2011	24 aprile
1974	14 aprile	1993	11 aprile	2012	8 aprile
1975	30 marzo	1994	3 aprile	2013	31 marzo
1976	18 aprile	1995	16 aprile	2014	20 aprile
1977	10 aprile	1996	7 aprile	2015	5 aprile
1978	26 marzo	1997	30 marzo	2016	27 marzo
1979	15 aprile	1998	12 aprile	2017	16 aprile
1980	6 aprile	1999	4 aprile	2018	1 aprile
1981	19 aprile	2000	23 aprile	2019	21 aprile
1982	11 aprile	2001	15 aprile	2020	12 aprile
1983	3 aprile	2002	31 marzo	2021	4 aprile
1984	22 aprile	2003	20 aprile	2022	17 aprile
1985	7 aprile	2004	11 aprile	2023	9 aprile
1986	30 marzo	2005	27 marzo	2024	31 marzo
1987	19 aprile	2006	16 aprile	2025	20 aprile
1988	3 aprile	2007	8 aprile	2026	5 aprile

Figura 2.4: Le date della Pasqua

L'effetto Pasqua, come è evidente, si cancella dopo un periodo di due mesi.

Per tener conto della Pasqua c'è la possibilità di introdurre una variabile che per i mesi di marzo e aprile è proporzionale al numero di giorni antecedenti e comprensivi della Pasqua che cadono nel mese. Entrambe le procedure effettuano in automatico un test preliminare sulla significatività di tali effetti.

2.2 Trasformazione dei dati

Molta della teoria statistica delle serie storiche poggia sull'ipotesi che i processi lineari si siano formati allineando nel tempo variabili casuali gaussiane o virtualmente tali. Nelle applicazioni reali la gaussianità è rara come l'unicorno e, sempre più spesso, si scelgono metodi robusti rispetto alle varianti della distribuzione marginale delle singole variabili casuali costituenti i processi.

Un modo, abbastanza diffuso e altrettanto discutibile, per imporre la gaussianità quando essa sia naturalmente assente è di procedere ad una traformazione dei valori della serie originale il cui istogramma sia più vicino alla gaussianità. Le trasformazioni più usate sono quella Box-Cox e quella di Mandel

$$Y_t = \begin{cases} \dfrac{(W_t - \beta_0)^\lambda - 1}{\lambda} & \text{per } \lambda \neq 0 \\ \log(W_t - \beta_0) & \text{per } \lambda = 0 \end{cases}; \qquad Y_t = \begin{cases} \dfrac{e^{\lambda W_t} - 1}{\lambda} & \text{per } \lambda \neq 0 \\ W_t & \text{per } \lambda = 0 \end{cases} \tag{2.4}$$

dove $W_t + \beta_0 = \min(W_t) + 1$.

La scelta di λ, per entrambe le trasformazioni, andrebbe guidata dalla applicazione di un qualche test che aiuti a controllare la maggiore gaussianità della serie trasformata. Di questi test ne esitono svariati, ma la loro applicazione non è sempre pacifica (si veda comunque il paragrafo 5.2.4 per un approfondimento). Una pratica semplice è quella di effettuare una scansione dei valori di λ, per esempio nel'intervallo $[-2.5, 2.5]$ con passo, diciamo di 0.1, e vedere per quale valore assumono le usuali misure di asimmetria e curtosi:

$$\gamma_1 = \frac{\sqrt{n}\sum_{t=1}^{n}(W_t - \bar{w})^3}{\left[\sum_{t=1}^{n}(W_t - \bar{w})^2\right]^{1.5}}; \quad \gamma_2 = \frac{n\sum_{t=1}^{n}(W_t - \bar{w})^4}{\left[\sum_{t=1}^{n}(W_t - \bar{w})^2\right]^2} - 3 \tag{2.5}$$

In particolare occorrerà valutare se si avvicinano al valore nullo caratteristico delle distribuzioni simmetriche e mesocurtiche tra cui ricade la curva di densità gaussiana. Un test che include questo ragionamento è quello proposto in Jarque e Bera (1980).

$$JB = \frac{n}{6}\left(\gamma_1^2 + \frac{\gamma_2^2}{4}\right) \tag{2.6}$$

L'istogramma, ovviamente, rimarrebbe invariato per qualunque riordinamento dei valori della serie storica e quindi non è uno strumento molto pertinente in questo ambito. Tuttavia, è in grado almeno di fornirci un'idea sulla asimmetria della distribuzione dei valori (campione di ampiezza uno da una serie storica di n valori che si assimila ad n campioni di ampiezza uno). Si ottengo inoltre indicazioni sulla curtosi nelle sue due valenze fondamentali: picco al centro e pesantezza delle code. Diciamo che in questo caso vale la pena di sacrificare la stretta applicabilità dello strumento d'analisi alla quantità di informazioni ottenute.

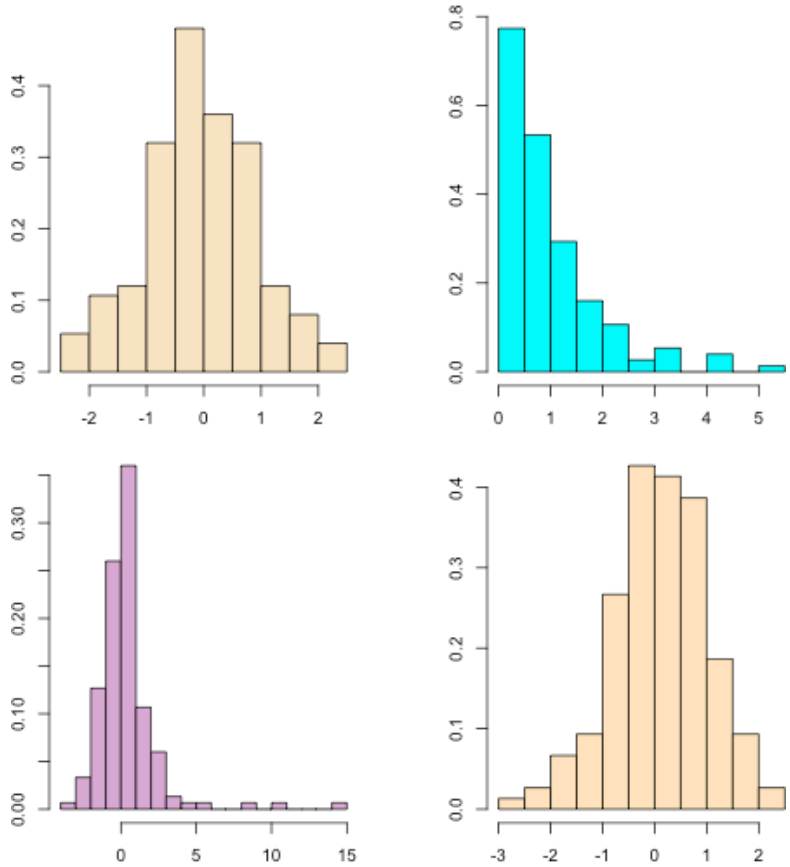


Figura 2.5: Simulazione di quattro serie storiche con marginali diverse

Nella figura 2.5 sono stati riprodotti gli istogrammi di quattro distribuzioni empiriche di una serie storica di $n = 150$ osservazioni provenienti da una gaussiana standardizzata, esponenziale, t-Student con due gradi di libertà e t-Student con 30 gradi di libertà. I risultati del test JB sono riportati nella tabella 2.1

Tabella 2.1: *Esempi per il test di Jarque-Bera*

Modello	JB	p-value
Gaussiano $(0, 1)$	0.3337	0.8462
Esponenziale $(\mu = 1)$	194.8452	0.0000
t-Student $\nu = 2$	2467.522	0.0000
t-Student $\nu = 30$	1.919349	0.3830

Il test fornisce indicazioni plausibili dato che ha un valor p molto basso per ditribuzioni manifestamente non gaussiane ed alto per la gaussiana o una molto simile quale la t-Student con 30 gradi di libertà.

Occorre dire che il presupposto in base al quale si effettuano le trasformazioni è che il tipo di processo che genera la serie storica rimanga tracciabile anche se questa subisce un cambiamento funzionale. Forse questo può essere vero per le tra-

sformazioni lineari, ma è dubbio per altre trasformazioni a forte nonlinearità come il logaritmo o l'esponenziazione. In fondo, l'unica vera ragione per avventurarsi nelle trasformazioni verso la gaussianità è la possibilità di proporre degli intervalli di previsione fondati su di una teoria solida e conosciuta. Anche su questo il dibattito è aperto: analoghi intervalli possono essere ottenuti con distribuzioni non-necessaiamente gaussiane (che però includono la gaussiana) che sono più robuste rispetto al comportamento selvaggio delle serie storiche del mondo reale.

2.2.1 Trasformazioni contro l'eteroschedasticità

La condizione che la varianza sia e si mantenga indipendente dal tempo è un altro mito che circonda l'applicazione della metodologia BJ. Infatti, come sarà ripetuto più volte, è raro che la omochedasticità temporale si verifichi nella realtà ed infatti esistono modelli che tengono conto in modo esplicito della sua violazione.

La suddivisione della serie storica in un certo numero di sottoserie sequenziali può consentire di cogliere la eteroschedasticità dei valori.

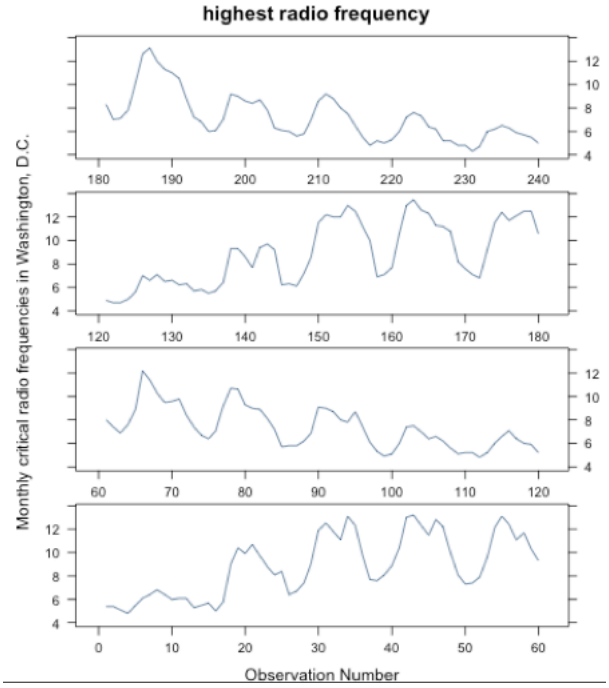


Figura 2.6: Frazionamento in quattro sotto serie

Un test datato, ma utile per verificare la stabilità della varianza o, almeno la sua esistenza finita è di realizzare un grafico in cui si pone sulle ascisse il numero di osservazioni della serie che si considerano e sulle ordinate la loro varianza campionaria. Nella figura 2.7 sono rappresentate due serie storiche: prezzi del grano plurisecolari e un processo formato da osservazioni simulate da una variabile casuale t-Student con 2gradi di libertà. Come è noto, in questo caso, la loro varianza è infinita.

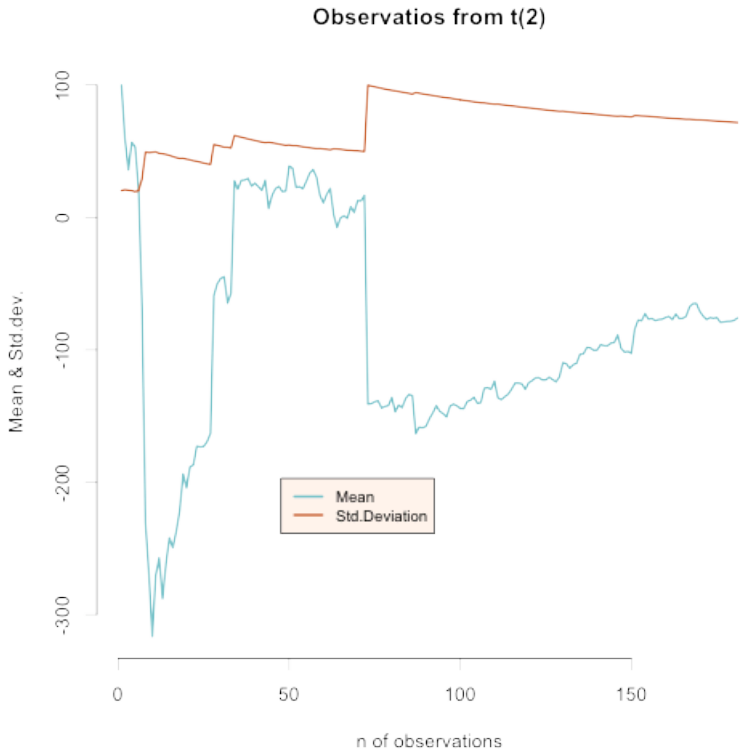


Figura 2.7: Verifica grafica della stazionarietà in media ed in varianza

Da notare che abbiamo casualizzato la sequenza dei valori per evitare che il grafico fosse compromesso dalla correlazione lineare tra osservazioni, tipica delle serie storiche. La varianza od anche solo la media non stazionaria si evidenzia con l'andamento dei grafici vicino all'estemo superiore (100).

In questo paragrafo cercheremo solo di attenuare l'impatto della eteroschedasticità sulla possibilità di applicare la metodologia BJ. Supponiamo che la varianza di una serie storica sia $var(W_t) = cf(\mu_t)$ dove c è una costante positiva e μ_t varia con l'indice temporale. Sarebbe molto utile poter disporre di una trasformazione dei dati $T(\cdot)$ tale che $var[T(W_t)] = costante$. Ipotizziamo che $T(\cdot)$ sia continua, monotona (almeno a tratti) e derivabile nei punti in cui è continua. L'espansione in serie di Taylor intorno a μ_t fino al primo termine porta alla relazione:

$$T(W_t) \approx T(\mu_t) + T'(\mu_t)(W_t - \mu_t) \tag{2.7}$$

Se tale approssimazione fosse accettabile, la varianza di $T(W_t)$ potrebbe, a sua volta, essere approssimata con

$$V[T(W_t)] \approx [T'(\mu_t)]^2 V(W_t) = [T'(\mu_t)]^2 cf(\mu_t) \tag{2.8}$$

Se ora poniamo

$$T'(\mu_t) = \frac{1}{\sqrt{f(\mu_t)}} \Rightarrow T(\mu_t) = \int \frac{1}{\sqrt{f(\mu_t)}} d\mu_t \tag{2.9}$$

che, se si definisce $f(\mu_t) = x^{\lambda}$, ci porta alla Box-Cox. L'effetto della trasformazione dipenderà dalla qualità della approssimazione lineare.
Ad esempio, se un trend accentuato ha prodotto la contemporanea presenza di valori molto piccoli e molto grandi, converrà sostituire ai valori originali i loro logaritmi.

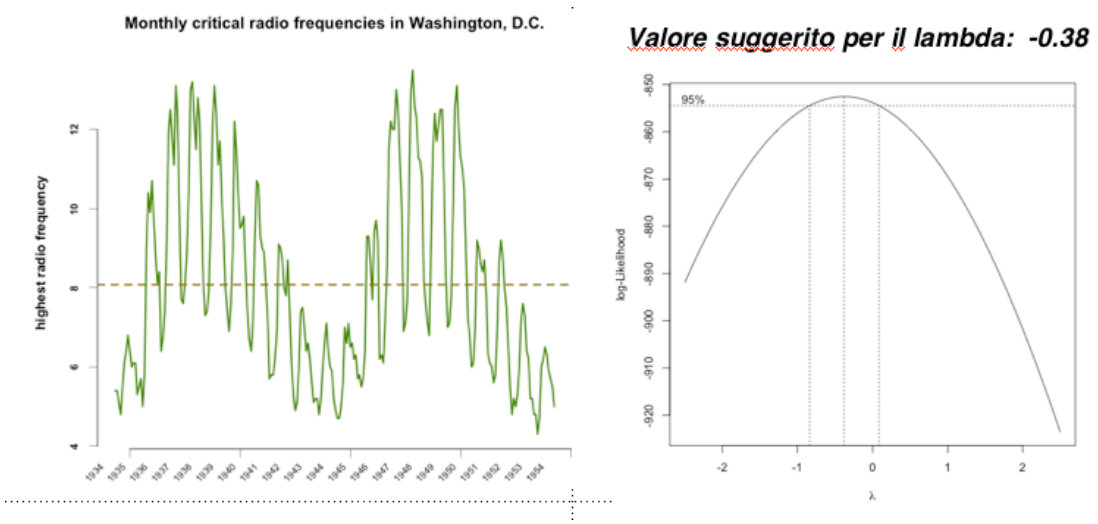


Figura 2.8: Caso in cui la trasformazione potrebbe risultare utile

La nostra impostazione vede la trasformazione Box-Cox come attività preliminare dell'analisi delle serie storiche. Non è l'unico modo di procedere. Una procedura alternativa vede la determinazione dei parametri della trasformazione insieme agli altri parametri dei modelli; è vero che i calcoli aumentano e si complicano, ma non possono considerarsi un problema, soprattutto se la procedura che vede ad esempio il λ della Box-Cox un parametro incognito, fornisce previsioni migliori.

La trasformazione logaritmica è quella più spesso effettuata nell'analisi delle serie storiche perché Il logaritmo ha effetto telescopico: ingrandisce le differenze piccole e riduce le grandi cosicché un trend esponenziale nei valori originari diventa un trend lineare nei logaritmi e quindi più facilmente eliminabile con le serie delle differenze.

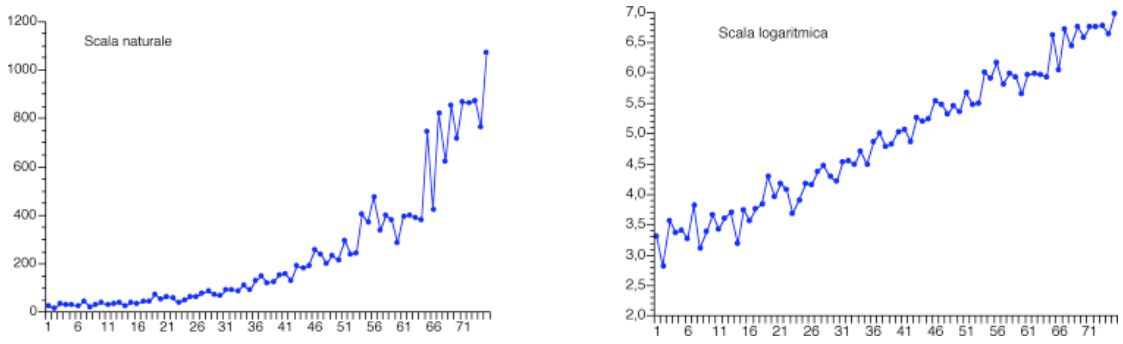


Figura 2.9: Effetto della scala logaritmica

Questo comportamento può essere sfruttato per presentare un andamento tormentato da picchi e valli con un profilo più addolcito o che comunque presenti la serie con oscillazioni più smorzate (quelle grandi percentualmente più toccate rispetto a quelle piccole) che porta in primo piano il gradiente del fenomeno.

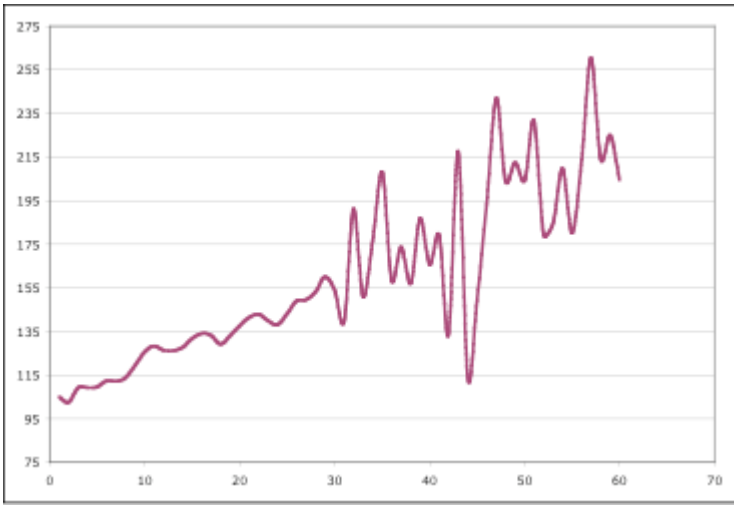


Figura 2.10: Esempio di trasformazione

Var(Y1:Y30)	235.269	0.015	Var(LY1:LY30)
Var(Y31:Y60)	788.472	0.016	Var(LY31:LY60)
Rapporto	3.351	1.120	Rapporto

Come mostrano i caloli associati alla figura 2.10, la scala logaritmica riduce la variabilità soprattutto tra i valori più grandi.

La trasformazione logaritmica elimina una particolare forma di nonstazionarietà in varianza. Si supponga che il processo sia costruito nel modo seguente: $W_t = a_t Y_t$ dove Y_t è un processo stazionario in media ed in varianza e gli a_t siano delle costanti che variano con il tempo ma rimanendo incorrelate con gli Y_t a qualsiasi ritardo. In particolare, si ipotizza:

$$E(Y_t) = \frac{1}{E(\mu_t)};\ V(Y_t) = \sigma_y^2;\ V(a_t) = \sigma_a^2;\ cov(a_t, Y_t) = 0 \tag{2.10}$$

La media e la varianza di W_t diventano allora $E(W_t) = 1$, $V(W_t) = \sigma_w^2 = a_t^2$ che, come si vede, dipende dal tempo t. L'uso della trasformazione logaritmica elimina questa particolare forma di eteroschedasticità. Infatti, ponendo $LW_t = \log(W_t)$ per ogni t allora LW_t non solo è stazionario in media come l'originale, ma lo è anche in varianza

$$\begin{aligned} V(LW_t) &= var[\log(W_t)] = var[\log(a_t Y_t)] && (2.11)\\ &= var[\log(a_t)] + var[\log(Y_t)] && (2.12)\\ &= \sigma_a^2 + \sigma_{LW}^2 \end{aligned}$$

che non dipende più da t. L'effetto dell'operazione è illustrato nella figura 2.11 in cui il grafico a destra si presenta con oscillazioni maggiormente smorzate rispetto alla serie storica a sinistra.

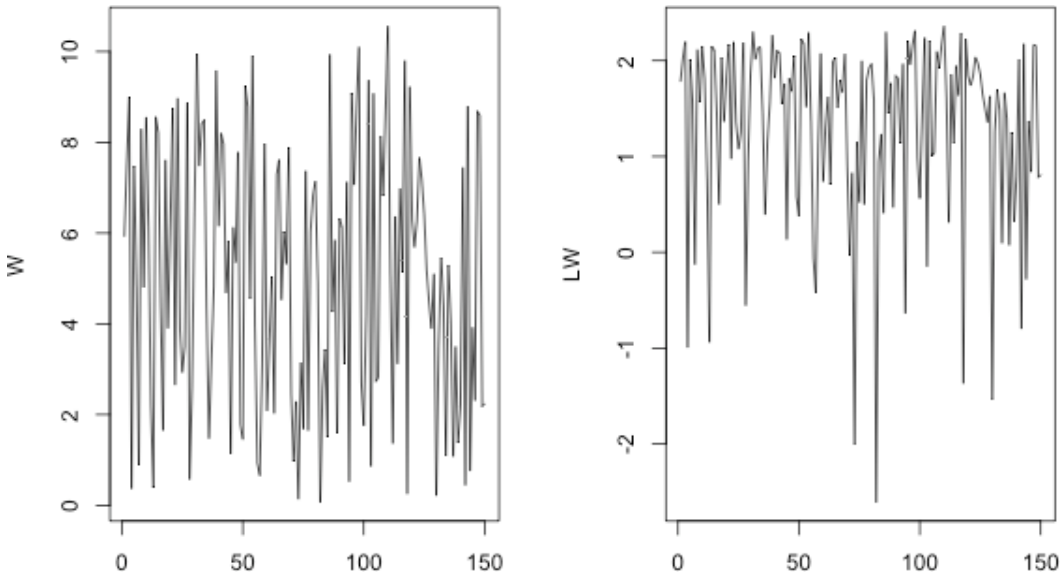


Figura 2.11: Effetto della trasformazione

L'efficacia della trasformazione logaritmica perdura fintanto che la eterogeneità nelle varianze è riconducibile alla struttura qui illustrata. Si è spesso notato un leggero miglioramento della procedura BJ applicata ai dati trasformati per la gaussianità e per l'omoschedasticità. Rimane però con l'incertezza della loro reale utilità. Infatti, ai fini delle previsioni sussiste la distorsione

$$E\left[\log\left(W_t\right)\right] \neq \log\left[E\left(W_t\right)\right] \tag{2.13}$$

e la distorsione (o *bias*) è maggiore per i valori grandi. Questo risultato è una conseguenza della disuguaglianza di Jensen (dal nome del matematico danese Johan Jensen) è una relazione che lega il valore di una funzione convessa al valore della medesima funzione calcolata nel valor medio del suo argomento: $f\left(E\left[X\right]\right) \leq E\left[f\left(X\right)\right]$

Un trucco empirico, utile per ridurre la distorsione nelle trasformate Box-Cox dei valori previsti è la seguente:

$$\bar{W}_t = \tilde{W}_t \left(\frac{GM}{AM}\right)^{\tau} \tag{2.14}$$

dove GM e AM sono, rispettivamente la media geometrica e la media artimetica della serie storica nella scala originale e τ è una costante da determinare tramite prove ed errori. Per un dato valore di τ, il fattore $(GM/AM)^{\tau}$ determina una misura di variabilità relativa che si mantiene sempre inferiore o al massimo uguale all'unità. Il fattore in (2.14) è una funzione monotona decrescente dell'esponente: maggiore è

τ più piccolo è il fattore e, di conseguenza, maggiore è il fattore di correzione (cioè la riduzione del valore in scala trasformata). per $\tau = 1$, expression (2.14) rende uguali la media aritmetica nella scala trasformata alla media geometrica nella scala originale.
Si dimostra che il rapportoGM/AMè uno stimatore asintoticamente efficiente della dispersione relativa di un campione ottenuto dalla distribuzione Pearson 3 di cui si riportano alcuni esempi nella figura 2.12

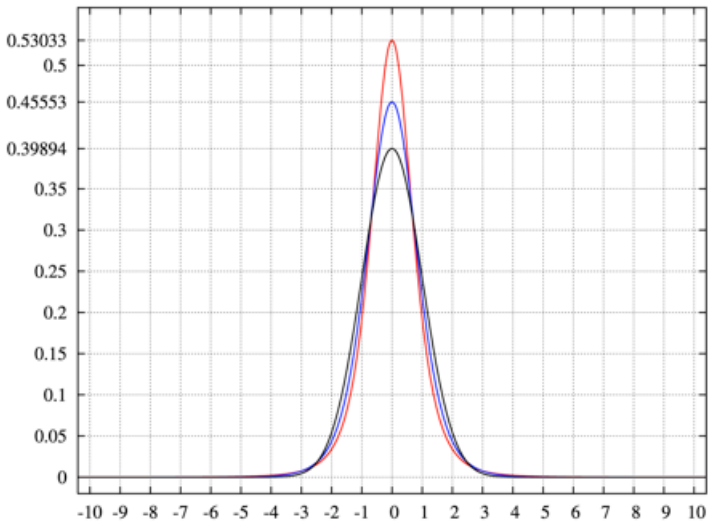


Figura 2.12: Alcuni esempi della distribuzione Pearson III

Per $\tau = 2$, la equalizzazione delle medie coinvolge la media aritmetica nella scala trasformata e la media armonica nella scala originale. In questo senso è bene ricordare che si ha $(AM)(HM) = GM^2$ purché $\log(y)$ abbia una distribuzione simmetrica, quale in fondo è l'obiettivo della trasformazione Box-Cox). Se $\tau = 0$ non si applica aln fattor di correzione dato che tale fattore moltiplicativo sarebbe uguale ad uno. Per $\tau < 0$ la media dei valori ri-trasformati sono aumentati dato che $AM \geq GM$.

2.3 Valori anomali nella serie storica

Lavorare con dati veri porta a trattare con dei valori irragionevoli ovvero dati molto discosti da tutti gli altri che si configurano come interrutivi del flusso "naturale della serie storica. E' chiaro che l'anomalia deve essere sempre provata e non può essere presunta in base a considerazioni intuitive o esoteriche. Per intenderci, questi valori strani o inusuali preferiamo definirli valori anomali o con il termine latino *remota* o anglosassone *outliers*.

Ogni volta che succede di incontrare un dato che desta sorpresa e/o allarme rispetto al resto dei dati e rispetto a quello che è ben consolidato nel sapere relativo al fenomeno che si studia, è necessario decidere se uno o più di questi *remota* siano dovuti a degli errori grossolani oppure a dissonanze tra osservatori diversi od ancora ad unità eccezionali od anche estranee chissà come finite nella nostra serie storica. In questi casi si rimane sempre perplessi e ci si domanda se si tratti di dati realmente pertinenti il fenomeno da cui dovrebbero essere scaturiti (e quindi confermarne la validità) o, in alternativa, di valori acquisiti in modo sbagliato e/o dalla fonte sbagliata e/o risultato di frode o sul quale manchi l'accordo tra gli utilizzatori potenziali. E qui che si deve valutare l'opzione di ignorarli, tenuto conto del rischio di escludere informazioni in verità genuine oppure elaborare anche dei dati sospetti, ma riducendone l'impatto per non danneggiare, eventualmente, la qualità complessiva dei risultati. La terza via, tenersi tutti i dati per esigenza di completezza e per avvantaggiarsi operando con una serie storica più grande e completa, può essere percorsa solo dopo che i dati siano stati controllati almeno per escludere errori manifesti.

C'è poi un'ipotesi suggestiva: che i *remota* siano in realtà un fenomeno latente che convive con quello che governa la casualità di ciò che studiamo e che solo di tanto in tanto dà un segno della sua presenza con dei supervalori (ad esempio, [**?**] propone questa interpretazione per i massimi delle portate dei corsi d'acqua). La congettura è fascinosa, ma di solito rimane tale poiché i supervalori non sono di solito presenti in misura adeguata per metterla alla prova.

Il senso comune può già aiutare nel riconoscimento di un dato anomalo. Nella tabella 2.2 sono riportate due serie storiche relative al traffico di persone e merci osservati per un periodo di 12 mesi. I grafici che le riguardano sono nella figura 2.13.

Tabella 2.2: Due serie storiche brevi

Mese	1	2	3	4	5	6	7	8	9	10	11	12
Persone (migliaia)	65	69	75	78	80	81	85	130	60	72	84	71
Colli (megatonn)	160	166	170	172	180	182	195	230	165	174	178	184

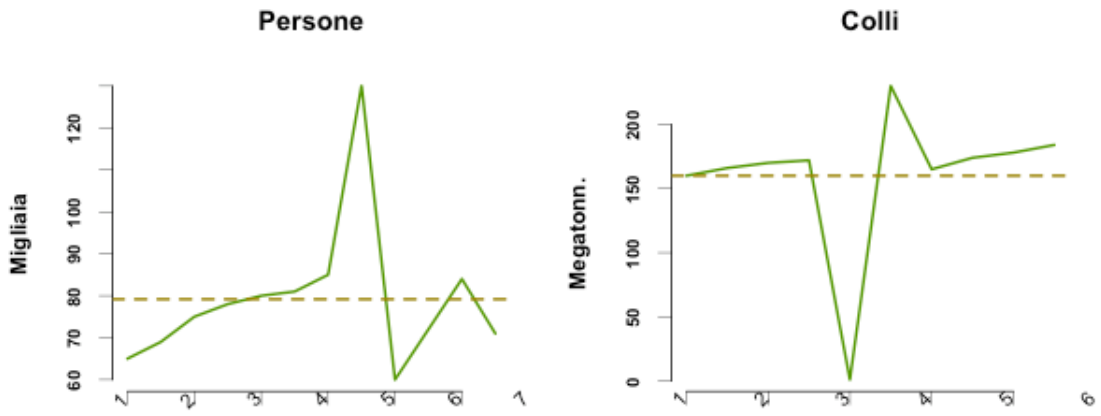


Figura 2.13: Anomalie nelle serie storiche brevi

Un conteggio $130mg$, non è tale da suscitare meraviglia in molti contesti, anche se è il doppio della rilevazione più piccola riportata nella riga; comunque, in diverse serie storiche con un *trend* molto marcato non è raro avere questi riscontri. Nella serie successiva del peso dei colli si pone il problema di credibilità per il dato: $230mt$ di peso. Tale valore è inusuale nell'esperienza -limitata- dei dati osservati, anche se, ora il valore sospetto è solo il 44% in più del peso minimo rilevato nella serie. Quando l'arco temporale di rilevazione è breve sarebbe presuntuoso pronunciarsi sull'anomalia perché altri nuovi dati potrebbero smentire il giudizio già dato.

Le anomalie nelle serie storiche spesso si presentano in coppie o in gruppi.

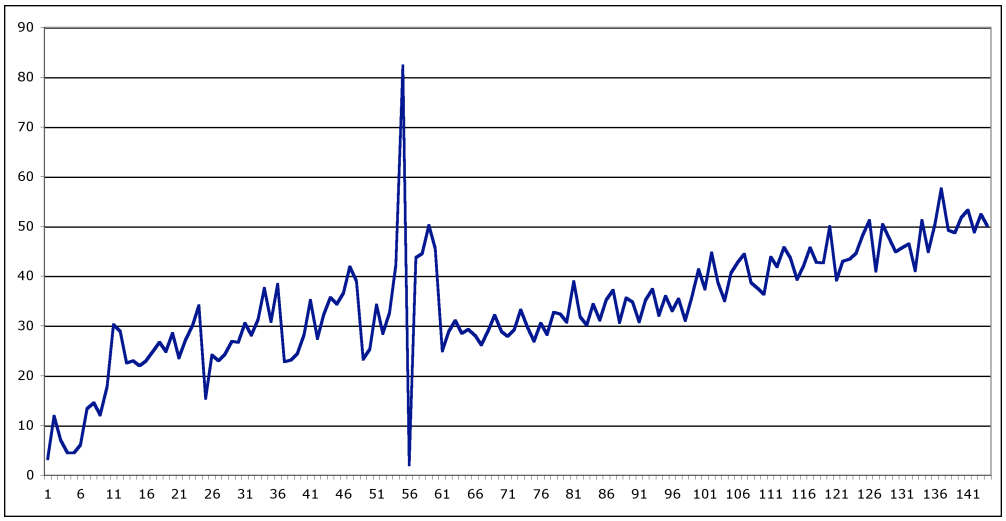


Figura 2.14: Valori anomali in fenomeni evolutivi

Nella figura 2.14 Un valore anomalo in eccesso è seguito da un valore anomalo in difetto a causa dell'attrazione del *trend* che ha come un effetto elastico.

In varie occasioni i valori anomali si presentano addirittura a "pacchetti" perché lo shock si estende in più periodi finché non interviene l'attrazione verso il *trend*.

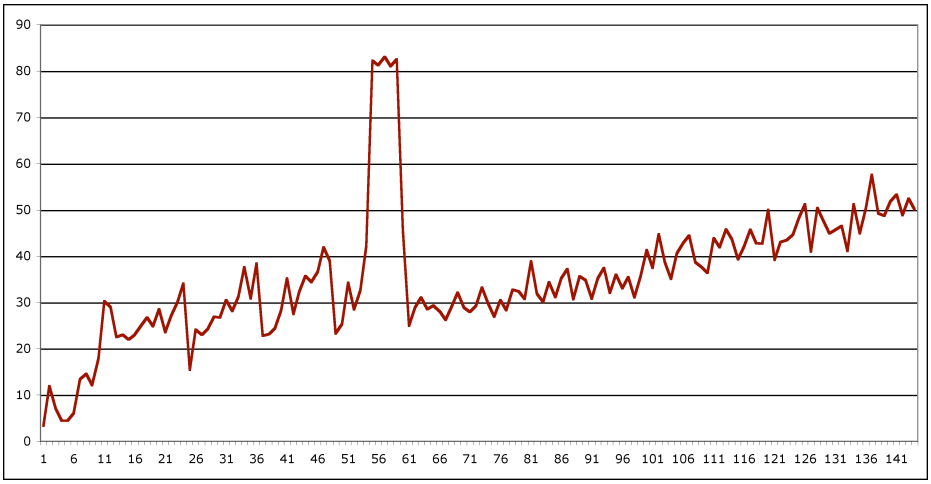


Figura 2.15: Valori anomali in fenomeni evolutivi

E' importante comprendere che i *remota* non sono per forza sbagliati o estranei ai dati o addirittura truffaldini; possono invece essere molto utili: ciò che per qualcuno è solo un rumore, per altri può essere un segnale importante.

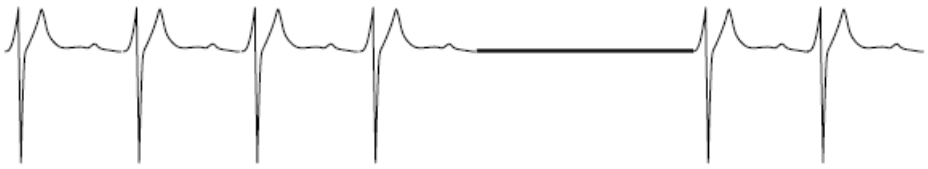


Figura 2.16: Tracciato di un elettrocardogramma

L'anomalia in questo caso è costituita dal *plateau* al centro. Il valore osservato non è anomalo in sé, ma il suo perdurare per un periodo così lungo ingenera preoccupazione.

Ogni valore anomalo è un valore estremo, ma non tutti i valori estremi sono frutto di anomalie. I valori anomali sono trattati con sospetto e fastidio dato che compromettono la rappresentazione della serie storica con un processo stocastico semplice e ben strutturato; c'è il rischio che una minuscola frazione dei dati finisca con il determinare l'interruzione di un'indagine oppure indurre nella scorciatoria di sostituire dei valori validi ma selvatici con altri più docili. Il fatto è che, in ogni insieme finito di n dati metrici, è sempre possibile riscontrarne un certo numero che si allontana dagli altri: ci sono sempre e comunque un valore minimo ed uno massimo e qualche altro valore che si avvicina all'uno o all'altro più di tutti gli altri. Se la separazione delle code dal resto della distribuzione è accentuata cominceremo a pensare ad una disfunzione nel fenomeno studiato: alcuni dati appaiono troppo diversi dagli altri. Non per questo si deve pensare ad errori o smottamenti nei

dati. L'evento raro esiste: ne siamo sorpresi, ma rientra comunque nel quadro del fenomeno: una temperatura di $-10°$ in una città Mediterranea, uno scarto del 25% di un prezzo, un patrimonio di molti milioni di euro per un disoccupato. Sono poco verosimili, ma legittimi fino a prova contraria e tanto deve bastare per trattarli come gli altri dati. Secondo la casa d'aste Christie's *non esiste una valutazione eccessivo, qualunque sia l'oggetto posto all'incanto.*

Nella definizione di valore remoto si annida una valenza di sorpresa; cioè, il sospetto su di un dato nasce perché non ci si aspettava di rinvenirlo di questa entità. Sul dato sospetto arriverà la luce dell'indagatore solo se lo si ritiene sorprendente anche rispetto a quello che risulta da osservazioni fatte in precedenza o in casi analoghi e/o che ci si poteva aspettare come allontanamento dal centro, nella teoria che circonda il fenomeno. L'affermazione che dei dati sembrano "incoerenti" è tanto cruciale quanto soggettiva e rappresenta l'ostacolo maggiore alla diffusione delle tecniche di riconoscimento e gestione dei dati fuori posto. E' importante poter decidere se il *remotum* è un elemento genuino della serie storica oppure è un contaminante che scaturisce da un fenomeno diverso da quelle studiato che si è infiltrato per errore o per altri motivi inconfessabili.

Un caso del genere sono le rarità cioè quei valori che esprimono gli aspetti meno frequenti e che possono apparire in forma più nitida solo con ampiezze campionarie consistenti: è la presenza di tali valori infrequenti, ma validissimi che rende problematico lo studio dei *remota*. Il nostro concetto di anomalia è di un valore che ha subìto una traslazione/espansione non prevista e sconosciuta che ne ha alterato la coerenza con gli altri spezzando la linea evolutiva attesa in base a quello che si vede prima e dopo l'evento. In effetti essi si distinguono rispetto agli altri destando allarme e sospetto o speranza.

Esempio 1. *I massimi mensili (ampiezza di 1 dml) delle precipitazioni giornaliere in una stazione delle isole Fiji sono riassunte nella tabella che segue.*

X_i	n_i				
0	1	39	10	11	7
1	2	47	11	12	7
2	3	43	12	13	7
3	4	43	13	14	4
4	5	40	14	15	5
5	6	41	15	16	2
6	7	28	16	17	4
7	8	20	17	18	2
8	9	16	18	19	0
9	10	11	>19		5

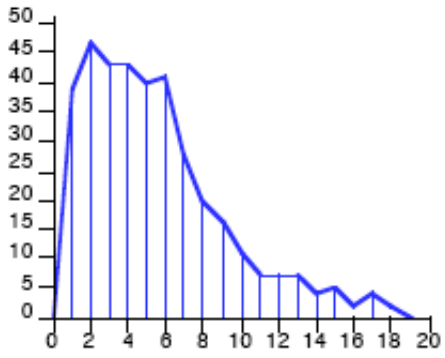


Figura 2.17: Valori anomali in fenomeni a coda pesante

Il diavolo si annida nelle code. In particolare, le code allungate possono dare luogo a valori anomali, anzi sono i valori anomali che talvolta provocano il distendersi delle code.

2.3.1 Riflessioni sui valori anomali

La facilità con cui si producono nuove informazioni e la riduzione dei costi di archivio ha portato alla costituzione diffusa e generalizzata di basi di dati enormi e di banche dati sugli argomenti più svariati. Un aspetto importante della interazione con le banche dati è di poter stabilire la parte dei dati compatibile con le condizioni di funzionamento ordinario dei modelli e degli schemi in cui i dati si sono inquadrati. In breve, è essenziale la ricerca ed il trattamento dei valori anomali o remoti.

I valori remoti hanno molte cause. Anscombe divide i *remota* in due categorie principali: quelli provenienti da errori nei dati e quelli che sono da attribuire alla variabilità intrinseca dell'aspetto a cui tutti i valori osservati fanno riferimento. Alcuni ricercatori sono convinti che si possa discutere di dato remoto solo quando è possibile fornire una spiegazione al suo accadere. Non dobbiamo dimenticare infatti che chi deve acquisire e controllare la qualità dei dati è continuamente costretto a prendere decisioni su di essi, giudicandoli validi o spurii e poi valutare se inserirli, non inserirli o includerli parzialmente. Tuttavia, alcune osservazioni possono risultare così lontane dalle altre da avere una influenza eccessiva sulla stima del modello di formazione del dato. Nelle applicazioni qualche dato scaturisce da circostanze inusuali: catastrofi, questioni internazionali, cambi di governo, scioperi o serrate, etc. C'è poi il rischio che certi dati siano sbagliati per mero errore materiale.

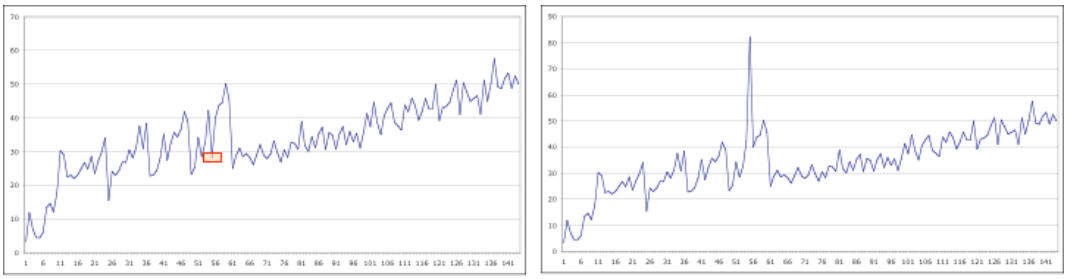


Figura 2.18: Tracciato di un elettrocardogramma

Il dato evidenziato nella figura 2.18 è 28.23, ma è stato inserito come 82.23 nel secondo grafico mostrando un picco fasullo. Brewer *et al.* raccolsero dei dati sulla paga oraria di un campione di infermiere. All'epoca, il salario si aggirava sui 12$ con una deviazione standard di 2$. Nel campione si trovò un'intervistata che dichiarava una paga di 42′000$. Si tratta di un chiaro errore umano: o chi risponde (od anche chi domanda) si riferisce all'importo annuale e non a quello orario. Qui la correzione è semplice: si divide l'importo annuale per le settimane e le ore lavorate, in media, per arrivare al dato orario.

Nell'esame di un qualsiasi campione di dati si incontrano talvolta campi con modalità o valori fuori le righe quali ad esempio temperature corporee di 73 gradi che sarebbero eccessive per una persona se fossero misurate in gradi Celsius e troppo basse se espresse in gradi Farheneit. È facile sospettare una inversione di cifre (37 e non 73 nei gradi Celsius) come spesso succede se i dati sono inseriti manualmente o con uno *scanner*. Potrebbe anche trattarsi di una carenza nella calibrazione, taratura, funzionamento degli strumenti, ma poiché non si è sicuri della validità della misurazione è più prudente trattarla separatamente se la misurazione non può essere ripetuta e riaccertata in modo convincente.

Categorie e numeri che destano molta sorpresa potrebbero anche essere solo il risultato di un tentativo di frode cioè dell'inserimento intenzionale di un dato aggiuntivo sbagliato o della alterazione di un dato buono. Quindi, qualche valore strano non del tutto compatibile con il profilo complessivo dei dati, è di normale riscontro nelle indagini sul campo. Solo se questi casi estremi sono molto remoti rispetto alle modalità loro contigue ovvero remoti rispetto a quanto ragionevolmente ci si può attendere dalla teoria del fenomeno, nascerà il sospetto di malfunzionamenti: se una ASL ha storicamente richiesto il rimborso di un numero mensile di "parti cesarei con complicazioni" che oscilla tra i 20 ed i 30 e, un dato mese, richiede rimborsi per 120 di tali operazioni, questo non necessariamente è un fatto anomalo, ma potrebbe essere la spia di un cambiamento nel meccanismo dei rimborsi o nella gestione della ASL. Non sempre è possibile decidere con tranquillità. Se gli accertamenti sul dato sbagliato non evidenziano errori di misurazione o di rilevazione od anche di comunicazione oppure questi spiegano solo in parte l'anomalia, allora il dato può ancora essere escluso o trattato, ma non dimenticato perché potrebbe sia essere il frutto di una contaminazione che di una truffa alla nigeriana.

La disponibilità di dati sufficienti alla ricostruzione del valore esplicitamente sbagliato evita di doverlo escludere e consente di recuperare quelle informazioni buone contenute nella serie storica con solo una parte, magari piccola, di valori anomali. Queste solo le situazioni più semplici da gestire perché il valore sospetto è manifestamente illogico e, corregendolo, non si trascura alcun segmento importante della popolazione in studio.

L'anomalia di per sé non ha una caratterizzazione sempre negativa. I valori remoti sono lampi di luce scorti da lontano. A produrre il brillìo può essere un coccio di bottiglia oppure uno smeraldo grezzo; per capire di che si tratta occorre andare a vedere. I *remota* sono osservazioni che interrompono la struttura che sembra avere la maggior parte dei dati loro contigui.

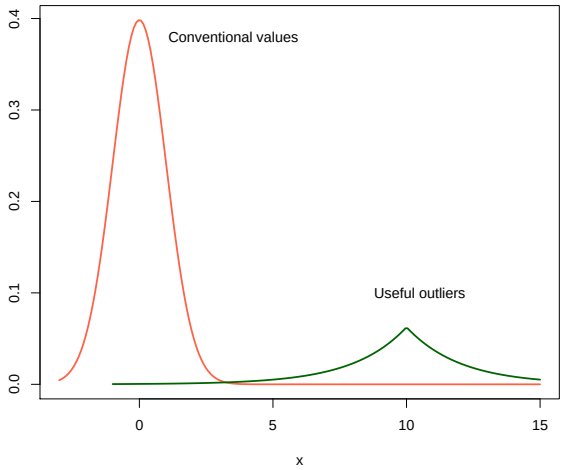


Figura 2.19: Scoperte nuove o peculiarità insospettate

Il fisico R. A. Millikan nel 1910 rigettò sei valori perché disomogenei rispetto agli altri. Uno di essi però presentava caratteristiche che meglio interpretate avrebbe rappresentato una delle prime prove dell'esistenza del *quark* che sebbene non abbia portato al premio Nobel può essere considerato uno dei risultati teorici più importanti ed utili del secolo scorso. Molte scoperte scientifiche si sviluppano intorno alla spiegazione di valori inattesi, dopo che verifiche accurate escludono semplici sbagli materiali. La figura 2.19 illustra bene il senso del nostro discorso: la grande maggioranza dei dati rientra nel perimetro dei valori convenzionali; nei gruppi isolati ci può essere la novità interessante.

La storia della scienza è ricca di episodi del genere. Ad esempio, l'argon fu scoperto nel 1894 da Lord Rayleigh e da William Ramsay (il primo ebbe nel 1904 il premio Nobel per al fisica, il secondo per la chimica) in un esperimento in cui esaminarono dei valori anomali di nitrogeno che non reagivano a delle sollecitazioni. Molte linee di ricerca nuove ed utili scaturiscono come serendipità piuttosto che da pianificazione. Darwin affermò che l'anomalia è il preludio di una nuova linea evolutiva. Alcuni ricercatori scoprirono in Africa delle donne che convivevano senza problemi con il virus dell'HIV per anni ed anni senza alcun particolare trattamento farmacologico laddove, la stragrande maggioranza, moriva purtroppo rapidamente. Se si agisse con leggerezza, i casi delle donne sopravviventi sarebbero scartati come valori anomali; invece debbono servire come ispirazione e suggerimento: qual è l'aspetto che le differenzia dalle altre? A quale caratteristica speciale, se c'è, può essere attribuita la resistenza al virus?

2.3.2 Individuazione dei valori anomali

Il principio che sottende ogni valutazione statistica dell'anomalia è che un valore remoto sia poco rilevante o dannoso per le elaborazioni in quanto ottenuto da fenomeni diversi o in condizioni diverse. I dubbi sui valori remoti possono essere risolti se il problema è suddiviso in due sottoproblemi. Innanzitutto, la questione della loro diagnostica in cui ci si preoccupa di individuare le osservazioni sospette e fornire un'idea della intensità della discordanza che esse presentano rispetto agli altri dati.

Il secondo problema concerne il trattamento dei valori remoti in modo che le conclusioni tratte dai dati non siano troppo influenzate dalla loro presenza o dalla loro esclusione. La prima argomentazione è la premessa logica della seconda: accertiamoci preliminarmente se una o più osservazioni sono disallineate rispetto al resto del campione o dell'insieme di dati e poi si vede quale possa essere la strategia più efficace per convivere con queste disgraziate.

Non tutti i valori sospetti sono da giudicare anomali e non tutti i valori anomali hanno sempre la cortesia di mostrarsi come sospetti (tendono a disperdersi nella normalità dei dati). È quindi importante avere un quadro delle possibili cause delle anomalie in modo da poter stabilire delle tecniche di diagnosi nonché delle procedure di trattamento dei casi eccezionali, una volta che se ne sia accertata la presenza e la necessità di intervento. L'ideale sarebbe un metodo che agisca selettivamente sugli elementi osservati nella serie storica: ne decreti la legittimità nel caso il valore sia genuino e ne stabilisca l'esclusione nel caso il valore sia estraneo. Questo maccanismo è solo ideale e non si può realizzare. Dobbiamo quindi adoperare tecniche che riducono l'impatto del valore sospetto sia quando è anomalo sia quando non lo è. Un meccanismo del genere è tanto migliore quanto più esalta il primo aspetto e attenua il secondo.

Un gruppo di tecniche semplici ed utili per la diagnostica dei valori anomali prevede la costruzione di intervalli di allerta o di pericolo che evidenziano il grado di sospetto associato ad una data osservazione. Nelle nostre indagini sulle serie storiche abbiamo distinto diversi metodi per costruire gli intervalli

- Soglie basate sui quantili

$$\begin{aligned}
\text{Pericolo inferiore} \quad & \frac{Q_1+Q_2}{2} - 3\,(Q_3-Q_1) \\
\text{Allerta inferiore} \quad & \frac{Q_1+Q_2}{2} - 1.5\,(Q_3-Q_1) \\
\text{Allerta superiore} \quad & \frac{Q_1+Q_2}{2} + 1.5\,(Q_3-Q_1) \\
\text{Pericolo superiore} \quad & \frac{Q_1+Q_2}{2} + 3\,(Q_3-Q_1)
\end{aligned} \tag{2.15}$$

- Soglie determinate in base a media e scarto quadratico medio sulla serie potata del $\alpha/2\%$ inferiore e superiore

$$\begin{aligned} \text{Pericolo inferiore} \quad & \mu_\alpha - 3\sigma_\alpha \\ \text{Allerta inferiore} \quad & \mu_\alpha - 1.5\sigma_\alpha \\ \text{Allerta superiore} \quad & \mu_\alpha + 1.5\sigma_\alpha \end{aligned} \tag{2.16}$$

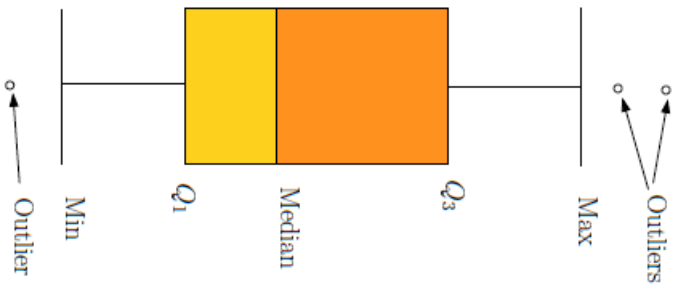


Figura 2.20: Boxplot per individuare gli *outliers*

L'idea guida è che i valori che superano le soglie di allerta debbano essere solo controllati per accertarne l'autenticità e l'assenza di errore mentre i valori che superano le soglie di pericolo dovrebbero essere trattati. E' chiaro che ogni definizione delle barriere produrrà decisioni legate a tale scelta; spesso, i valori di soglia scaturiscono da un preciso modello dei dati (ad esempio quello gaussiano) anche se non lo si rende evidente in modo esplicito.

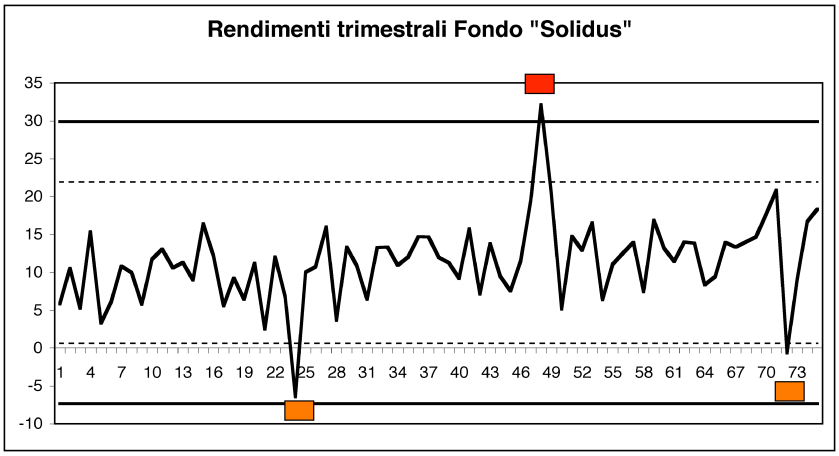


Figura 2.21: Rappresentazione delle soglie di anomalia

Ecco un esempio in cui sono stati inseriti deliberatamente dei valori anomali.

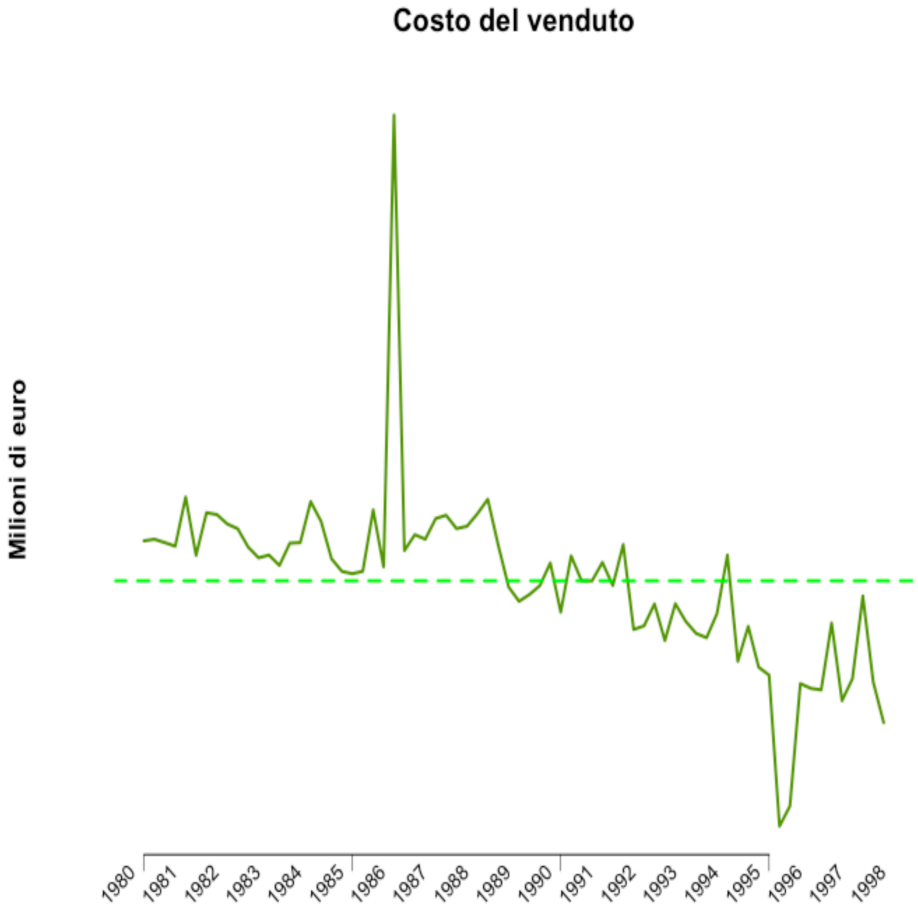


Figura 2.22: Soglie di anomalia e serie storiche

Il primo tipo di barriere è più attendibile, ma molto severo comportando molti controli; il secondo tipo è meno sensibile, ma si rischia di ignorare valori irragionevoli.

Barriere con mediana e diff. Interquart..

```
pericolo sup. frazione= 1   anno= 1986   valore=  490.02 pos: 25
pericolo inf. frazione= 2    anno= 1995   valore=  20.91 pos: 62
allerta inf. frazione= 3   anno= 1995   valore=  34.16 pos: 63
allerta inf. frazione= 4   anno= 1996   valore=  103.73 pos: 68
```

Barriere con media e sc.quadr.med.

```
pericolo sup. frazione= 1   anno= 1986   valore=  490.02 pos: 25
allerta inf. frazione= 2   anno= 1995   valore=  20.91 pos: 62
allerta inf. frazione= 3   anno= 1995   valore=  34.16 pos: 63
```

Figura 2.23: Calcolo delle soglie di anomalia

In questo caso il primo tipo di barriera risulta più efficace.

Esempio 2. *La serie D del testo di Box-Jenkins si riferisce alla lettura oraria della vischiosità di un processo chimico*

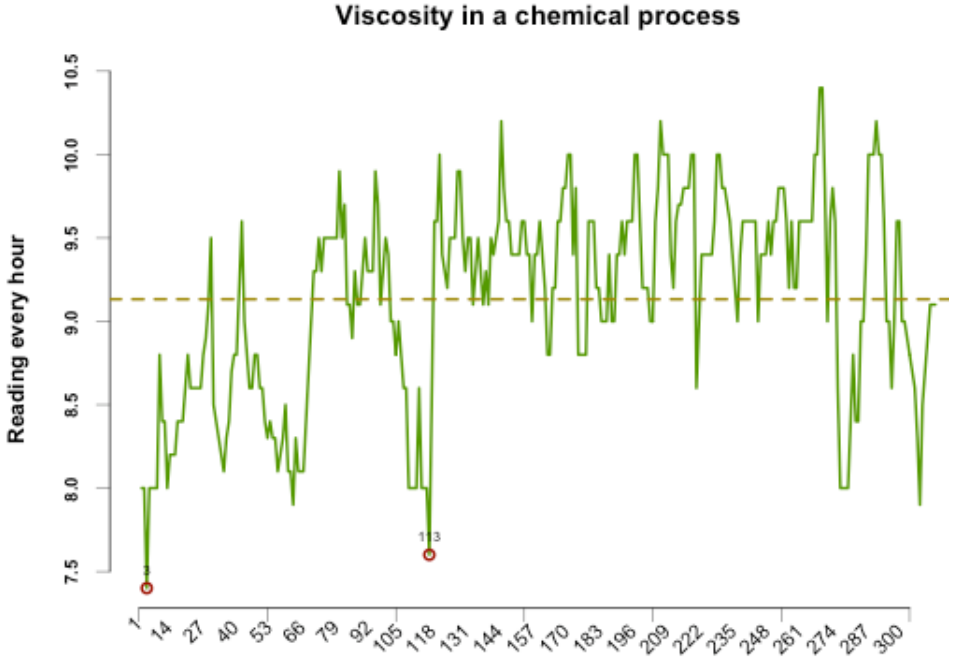


Figura 2.24: Serie D di Box-Jenkins. Due valori sospetti con $\alpha = 10\%$

Il valore sospetto è il 217°. Le soglie potate evidenziano invece il 3° ed il 113°.

Procedura di Basu-Meckesheimer

Si propone di proiettare il valore W_t della serie storica al tempo t in uno scenario dei valori precedenti e susseguenti, diciamo per τ valori che precedono ed altrettanti che seguono di modo che la sottosequenza che fa da scenario locale abbia lunghezza 2τ. Il confronto si realizza adoperando la quantità

$$b_t = \frac{|W_t - M_t|}{Q_3(W) - Q_1(W)} \qquad \tau + 1 < t < n - \tau \tag{2.17}$$

dove M_t è la mediana del vicinato $(W_{t-\tau}, W_{t-\tau+1}, \cdots, W_{t+\tau-1}, W_{t+\tau})$ ed in cui $Q_3(W) - Q_1(W)$ è la differenza interquartile dei valori inclusi nella intera serie storica.

Gli autori non chiariscono se il valore esaminato (e quindi sospettabile di anomalia) è incluso o escluso nel calcolo della mediana. D'altra parte, la formula (2.17) non effettua la ricerca di *outliers* nei primi e negli ultimi posti. Nella versione programma nel software del corso, il dato sotto osservazione è omesso e la disamina delle anomalie è estesa a tutta la serie.

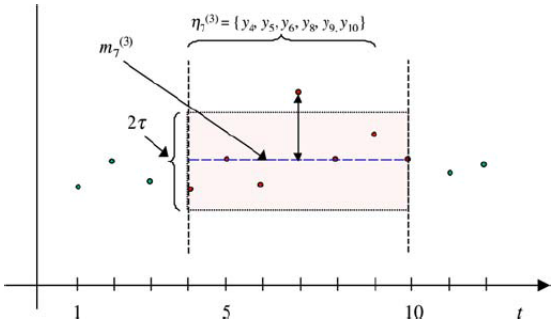


Figura 2.25: Metodo della mediana per la mondatura (*cleaning*) dei dati

Inoltre, la versione pubblicata contiene un altro parametro ovvero una soglia che aiuta a stabilire se un certo dato è un *remotum* oppure è da considerarsi normale. A nostro avviso questo complica la realizzazione della procedura. Il rapporto b_t è invariante rispetto alla scala di misurazione. I valori maggiori di b_t candidano le rispettive osservazioni all'anomalia. Tuttavia, così come è formulata, la procedura tende ad individuare anomalie che si presentano come picchi, locali o globali. Sembra invece ignorare le anomalie che si configurano come valli. La soluzione che abbiamo individuato è di ripetere la ricerca dei remota sulla serie trasformata come $Z_t = 1/W_t + 0.0000001, \qquad t = 1, 2, \ldots, n$ che copia la serie storica in uno specchio. I valori maggiori della somma dei $b_t\,(W)$ e dei $b_t\,(Z)$ indicheranno le anomalie.

Esempio 3. *Serie storiche dall'archivio della statistica italiana (Istat)*



La procedura, tra i primi tre sospetti evidenzia la coppia 1986/1987 come una combinazione picco-valle che richiede asseveramento dei dati e/o una spiegazione. L'anomalia al primo dato è determinato dal forte trend iniziale e quindi non deve preoccupare.

Un test dedicato

Il valore anomalo nelle serie storiche è tale quando si innalza o si abbassa in modo eccessivo rispetto ai valori che lo precedono e che lo seguono. L'evento t_2 nella figura 2.26 è pari all'evento in t_1, ma questo è del tutto normale nel suo contesto.

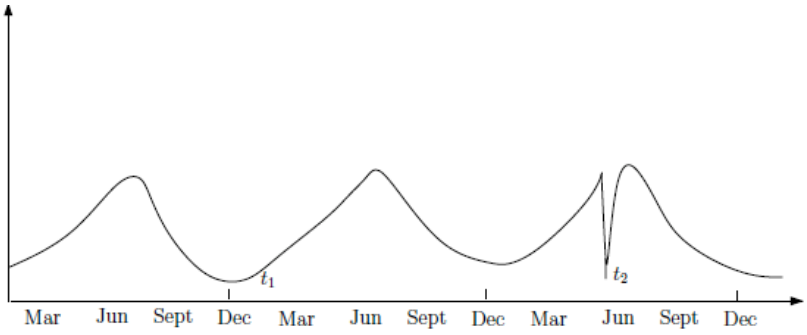


Figura 2.26: Valore anomalo contestuale

Un esempio di questo genere lo si può trovare nella ricerca di frodi negli acquisti con carta di credito. Una persona che spende in media 20 euro al distributore di benzina e 200 euro in abbigliamento non desterà alcun allarme se spenderà 20 in abbigliamento, ma allerterà i servizi di controllo del gestore della carta se spende 200 euro al distributore. La media dei due valori è uguale ed è uguale anche la varianza, ma i contesti sono differenti.

La 2.26 è il substrato per un test dedicato. Sia M il ritardo massimo per il quale sia consentita una stima almeno plausibile delle autocorrelazioni semplici; diciamo

$$M = \lfloor \sqrt{n} \rfloor \qquad \text{oppure} \qquad M = \lfloor n/4 \rfloor \tag{2.18}$$

Calcoliamo le autocorrelazioni semplici fino al *lag* M utilizzando l'intera serie storica $r_1, r_2, \cdots, r_M$. Per ogni osservazione t in $W_1, W_2, \cdots, W_n$ calcoliamo le autocorrelazioni semplici fino al lag M, della serie storica in cui però il t-esimo dato sia stato escluso: $r_1(t)^*, r_2(t)^*, \cdots, r_M(t)^*$. L'anomalia di un dato dovrebbe avere effetto sulle autocorrelazioni. Un modo per accorgersi della presenza di una osservazione abnorme è di confrontare i due correlogrammi e calcolare la loro distanza. Qui si propone di usare la metrica

$$d_t = \sqrt{1 - cor\left[r_i - r_i(t)^*\right]^2} \tag{2.19}$$

I candidati all'anomalia sono i valori che comportano la più grande distanza euclidea d_t. Le decisioni sull'anomalia andranno prese comunque con prudenza.

Esempio 4. *Una ricca fonte di serie storiche è reperibile nel sito* `http://www.stats.uwo.ca/faculty/aim/epubs/mhsets/readme-mhsets.html` *che contiene più di* 300 *serie distribuite in* 24 *cartelle. Nella serie riprodotta in figura sono stati inseriti due dati anomali puntualmente identificati dal criterio*

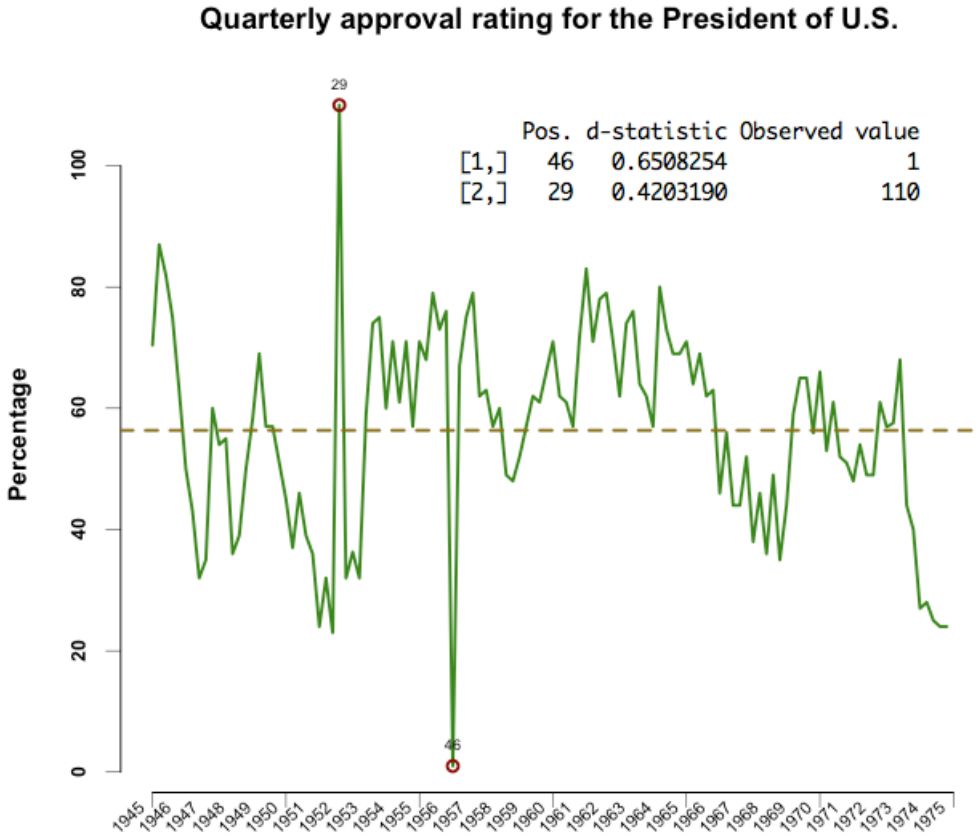


Figura 2.27: Calcolo delle soglie di anomalia

La diagnostica è interessante, ma ogni anomalia contestuale andrebbe esaminata attentamente prima di sottoporre il corrispondente dato a revisione.

I calcoli debbono però limitarsi alle sole coppie di osservazioni in cui entrambe hanno un valore valido e non mancante. Questo potrebbe comportare coefficienti di autocorrelazione calcolati su un numero di coppie meno numeroso di quanto possibile; una differenza che si aggiunge al degrado di attendibilità che già si riscostra per i *lag* più elevati.

2.3.3 Trattamento dei valori anomali

Eventi straordinari ed isolati nei processi stocastici non possono essere gestiti facilmente anche se ne compromettono seriamente l'andamento. Si rende necessario intervenire. Le strategie possibili sono tre di cui solo una richiede ulteriori approfondimenti

- Ci teniamo il dato così com'è.
- Effettuiamo delle verifiche e in caso di conferma della autenticità del valore correggiamo il dato sospetto sostituendolo con un altro -sintetico- più ragionevole.
- Utilizziamo procedure che non risentano dei dati anomali o che riducano al minimo l'effetto delle contaminazioni, reali o solo sospette che siano. Ad esempio, i modelli di intervento discussi da Box e Tiao (1985) che riacconciano efficacemente la serie storica colpita dalle anomalie

La reazione alla presenza degli *outliers* deve essere molto razionale. Siamo davanti ad un *trade-off*: escluderli, con il rischio di trascurare informazioni importanti oppure includerli, con il rischio di inficiare un calcolo o fuorviare la percezione di un profilo informativo. Non vi è dubbio che, qualora il valore spurio sia dovuto ad un errore di osservazione oppure ad un errore di registrazione i valori debbano essere riaquisiti e se questo non è più possibile o troppo costoso, debbono essere esclusi dalla serie storica, perché non ne possono logicamente e legittimamente far parte.

La verifica del protocollo di acquisizione dei dati è dunque indispensabile. Occorre eliminare il sospetto che aleggia su qualcuno dei valori della nostra serie: che siano stati eliminati per favorire o non ostacolare la conferma o la sconferma di una certa ipotesi. Se si riesce ad individuare l'errore allora si corregge e si immette il dato corretto, ma se non si riscontra alcuna causa di errore attribuibile al meccanismo di acquisizione, allora il dato deve essere considerato valido anche se questo significa compromettere tutte le stime e verifiche di ipotesi dell'indagine statistica. Magiore è l'isolamento o la eccezinalità del valore, maggiore è la probabilità che si tratti di un *remotum*,; ma se il dato è attendbile dal punto di vista delle procedure di ottenimento, allora diventa anch grande l'importanza da dare all'anomalia.

Il *follow-up*, quindi, racchiude tutti gli sforzi che si debbono produrre per poter escludere in modo inoppugnabile le osservazioni assolutamente fuori luogo. Insomma se il modello fallisce ciò deve poter essere attribuito esclusivamente al modello o alle procedure di analisi e non ai dati a cui sono state applicate. Ricordiamo che buttare via un qualsiasi dato altera in qualche modo l'immagine che la serie storica può dare del fenomeno in studio e si può procedere alla sua esclusione solo se c'è una ragione valida per farlo; molto spesso si tratta solo di valori alterati dalle condizioni sperimentali ovvero sono una spia che il processo di acquisizione è uscito dagli standard e si può rinunciare a loro senza troppe preoccupazioni (magari ripetendo le misurazioni ove possibile); tuttavia, adoperando tutte le osservazioni possiamo affermare con sicurezza che si è intenzionati a sfruttare del tutto il prelievo di informazioni realizzando con l'osservazione dei fenomeno.

La sostituzione dei valori anomali avviene con meccanismi identici alla imputazione dei valori mancanti discussa nel paragrafo che segue: il dato manifestamente anomalo è sostituito con un codice dipo "NA, cioè *not available* o non disponibile. La possibilità di coinvolgere tutti i valori, ma dando a ciascuno un peso attribuito in ragione della loro diversa attendibilità non sarà trattata in questo corso.

2.4 Valori mancanti

La qualità si riferisce alle caratteristiche intrinseche del prodotto e del processo produttivo. Per i dati statistici, la valutazione degli standard qualitativi avviene in base ad alcune caratteristiche, prima tra queste, l'attendibilità delle stime in termini di precisione, seguita dalle altre e cioè la tempestività, la comparabilità nel tempo e con altre fonti, la pertinenza rispetto ai bisogni degli utenti; oggi, poi, si fa sempre più attenzione al costo. Quindi, il nucleo principale di ogni approccio alla qualità dei dati statistici, che punti a dati attendibili, è costituito dall'esame delle singole fasi del loro processo produttivo. Tra esse, è importantissima la fase di acquisizione dei dati, che consente di valutare la qualità dei microdati, che costituiscono l'informazione basilare per la creazione dei macrodati. Da qui, l'attenzione all'esito delle risposte ai questionari delle indagini statistiche classiche, sia a quelli redatti su supporti cartacei, sia a quelli derivanti dall'uso di nuove tecnologie.

2.4.1 Ragioni dei valori mancanti

L'omissione di alcuni dati può essere dovuta a fattori procedurali non dipendenti dalla volontà dei ricercatori, come ad esempio errori umani, dati smarriti, dati cancellati erroneamente o a causa di un malfunzionamento dei sistemi di acquisizione, interruzione imprevista di un esperimento, modifica involontaria delle condizioni ambientali della rilevazione. Possono anche mancare dei dati a causa della cancellazione dovuta a inconsistenza con dati già registrati oppure dati non inseriti per incomprensione, dati non acquisiti perchè non considerati importanti al momento dell'inserimento; può inoltre succedere che alcuni dati vadano perduti inavvertitamente (o per deliberato intervento esterno) o che non siano rilevati a causa dei costi elevati di acquisizione.

L'assenza di una parte dei dati è un problema comune a molte indagini: tutti gli strumenti adottabili per la prevenzione di tale fenomeno possono solo ridurne l'intensità, ma non riescono in ogni caso ad eliminarne del tutto la presenza. Le mancate acquisizioni hanno due effetti sui risultati finali:

- riducono la quantità di informazione disponibile (nel caso di indagini campionarie, attraverso la riduzione della numerosità campionaria, viene prodotto un incremento del relativo errore di campionamento);
- comportano distorsioni nelle stime quando il meccanismo che le genera è in qualche modo collegato ai valori degli indicatori usati nelle analisi.

Il trattamento delle mancate risposte ha lo scopo di prevenire le distorsioni che la loro presenza può provocare sui risultati finali del processo di produzione dell'informazione statistica. Inoltre le mancate risposte provocano dei dilemmi in fase di analisi dei dati relativi alla maggiore difficoltà incontrata per effettuare le analisi sui dati incompleti. Questi, infatti, richiedono metodi complessi per la stima dei parametri che potrebbero non essere disponibili nei software statistici solitamente utilizzati per l'analisi dei dataset completi. Nella tabella che segue c'è un esempio sull'effetto dei valori mancanti in un *test* del confronto tra medie in cui si vede che l'esito risulta stravolto dall'impatto dei valori non riportati.

Tabella 2.3: Test t dei due campioni (test di Welch)

Con valori mancanti $Salario_var$ secondo il Genere	Senza valori mancanti $Salario_var$ secondo il Genere
t = 2.2434, df = 13.047	t = 4.6753, df = 19.476
p-value = 0.04286 $H_1 : m_1 \neq m_0$	p-value = 0.0001552 $H_1 : m_1 \neq m_0$
99 C.I.: -77.69913 532.16341 sample estimates: Media $gruppo_0$ media $gruppo_1$ 1716.875 1489.643	99 C.I.: 117.0161 483.5077 sample estimates: Media $gruppo_0$ media $gruppo_1$ 1733.929 1433.667

L'assenza di un valore può avere più significati:

- Il valore "esiste, ma può non essere noto" (es. l'età di una persona)
- Non si sa se il valore "esiste o meno" (es. reddito della persona)
- Il valore "è assente perchè non può esistere" (ad es. il numero di gravidanze di un maschio)

La conoscenza dei principali fenomeni socio-economici risulta indissolubilmente legata alla collaborazione delle unità di rilevazione, e ciò è tanto più vero quando il fabbisogno informativo non sia limitato a qualche dato aggregato, ma voglia sceverare gli aspetti più reconditi e riservati del comportamento individuale. Ne consegue che un limite tipico alla conoscenza della realtà fenomenica è quello imposto dalla reticenza dell'unità di rilevazione che si traduce nel rifiuto a collaborare, in tutto o in parte, agli obiettivi della ricerca, ovvero nella dissimulazione dell'informazione

fornita. Le mancate risposte ricorrono ogni qualvolta un dato potenzialmente osservabile viene a mancare costituendo delle vere e proprie discontinuità del processo di rilevazione che possono avere rilevanti effetti sulle statistiche campionarie.

2.4.2 Mancate risposte parziali

La problematica relativa alle mancate risposte parziali presenta peculiarità tali da conferirle un certo grado di autonomia rispetto a quella delle mancate risposte totali: la più apparente è la disponibilità di informazioni sul nonrispondente che, nelle indagini condotte su larga scala, può essere relativamente elevata per i nonrispondenti parziali, soprattutto per ciò che concerne alcune caratteristiche demografiche e socio-economiche. D'altronde, esiste qualche buona ragione per ipotizzare che il meccanismo generatore delle mancate risposte parziali sia qualitativamente diverso da quello relativo alle mancate risposte totali: nel primo caso l'unità di rilevazione è già stata contattata ed ha mostrato la sua disponibilità a collaborare, per cui la prima fase decisionale (relativa alla partecipazione tout court all'indagine) è stata superata favorevolmente; il nonrispondente parziale è infatti colui che avendo già deciso di prender parte alla rilevazione non sa o non vuole rispondere ad alcuni quesiti o fornisce una informazione manifestamente errata. Anche sotto il profilo eziologico la distinzione tra le due tipologie di nonresponse non è priva di contenuto; i fattori che comunemente concorrono alla determinazione di un insieme di dati "parzialmente" incompleto sono:

- rifiuto a rispondere;
- incapacità a fornire la risposta;
- soggetto rispondente errato.

Il primo fattore è quello di maggior interesse, dal momento che presenta caratteri di sistematicità, andando a colpire informazioni riservate o che l'individuo giudica non esternabili; è tra l'altro uno dei fattori meno controllabili da parte del rilevatore. Ad esempio, è opinione comune che le mancate risposte ad un quesito sul reddito siano da ascrivere prevalentemente a questo fattore, anche se non è assurdo ipotizzare che nei casi di occupazione precaria e negli strati sociali più poveri entri in gioco l'ignoranza del dato. Non è nemmeno da trascurare la possibilità che il dato sia ampliato per apparire meno povero. Il passo successivo consiste nell'individuare le determinanti del rifiuto: ad esempio, l'item nonresponse al quesito sul reddito è quasi sistematicamente accreditato nelle trattazioni specialistiche come uno dei casi in cui la propensione a rispondere è correlata al valore della caratteristica medesima, nel senso che sarebbe tanto più bassa quanto più alto il reddito. La situazione non è così

disperata come potrebbe apparire a prima vista dal momento che le informazioni disponibili consentono di far luce su alcuni aspetti del meccanismo generatore delle mancate risposte. In particolare, è possibile valutare il ruolo del livello di istruzione, dell'età, della condizione professionale, della numerosità della famiglia, dell'ampiezza demografica del comune di residenza e a volte anche di alcune importanti covariate del reddito, perchè i relativi dati sono generalmente disponibili su tutte le unità del campione.

2.4.3 Dati mancanti perchè impossibili

In questa categoria rientrano i dati che sono non rilevabili tenuto conto del tipo di unità cui si riferiscono. Ad esempio le domande riguardanti l'anno del matrimonio o del divorzio per i soggetti che non sono mai stati sposati o non hanno ancora divorziato. Più che mancanti si tratta di dati impossibilitati ad esistere quale ad esempio la regione di residenza di un immigrato o una immigrata non ancora usciti dalla clandestinità; ne consegue che per questa tipologia di dati mancanti non si dovrà procedere a nessun tipo di ricostruzione o di recupero dei valori assenti ma rimarranno come legittime lacune all'interno del dataset.

2.4.4 Dati volutamente omessi

Talvolta il dataset con cui si lavora contiene osservazioni molto rare o inattese nel suo contesto applicativo. Di per se non destano particolare allarme fin tanto che rimangono si unici nei record acquisiti, ma legittimi almeno dal punto di vista della rilevazione, se però ad una valutazione meno superficiale questi dati si rivelassero platealmente assurdi, perchè il dato fornito viola palesemente un vincolo logico-matematico, macroscopicamente irragionevoli, o addirittura errori di misurazione grossolani si aprirebbe la questione se mantenerli nel loro contenuto originale oppure accantonarli considerandoli come mancanti. Il principio dello sfruttamento massimo dei dati acquisiti spesso con sforzi e costi non indifferenti trova la sua origine nella qualità complessiva dei dati che devono essere depurati da contaminazioni imprevedibili e, ancora peggio, da frodi o manomissioni deliberate dei dati. A questo fine riteniamo che, dopo aver fatto tutti gli sforzi per accettare l'autenticità dei dati molto sospetti, se permangono ancora dubbi, sia meglio escluderli dalle elaborazioni considerandoli come veri e propri dati mancanti. Con questo si intende dire che le rilevazioni statistiche completate con i metodi di imputazione possono avere solo una ridottissima applicazione per finalità amministrative. Per esempio, se fosse il Ministero delle Finanze a richiedere i dati sugli addetti per comminare una sanzione

Tabella 2.4: Esempi di valori manifestatamente illogici

Salario-var	Anni-istruz	Età lavor	Anni pres	Corsi form
1485	NA	31	7	5
NA	10	**12**	**75**	3
1260	9	22	NA	4
1465	11	NA	11	3
1720	14	NA	16	1

ovvero imporre un certo livello di tassazione, difficilmente potrebbe essere soddisfatto con dei dati anche solo in minima parte artificiali; anche perchè l'impresa potrebbe opporsi con buone ragioni ed aprire un lungo contenzioso. Diverso, invece, è il caso di una società che intende conoscere la consistenza in un'area delle imprese di una certa dimensione. Infatti nei primi due esempi c'è un risvolto giuridico per cui occorre garantire all'impresa di essere conosciuta per i dati esatti che la riguardano. Nell'ultimo esempio la completezza rigorosa dei dati e delle informazioni non è indispensabile. Non bisogna trascurare che in alcuni casi il dato è mancante per ragioni legittime, quali la non applicabilità o la non pertinenza delle richieste o, addirittura, l'obbligo di astensione del rispondente. In queste situazioni la ricostruzione di valori omessi è una ingerenza forte nel *data set* che può rivelarsi controproducente.

2.5 Strutture possibili dei dati mancanti

Ogni analisi effettuata su dati incompleti richiede la specificazione sia della distribuzione di probabilità (modello) dei dati, sia della distribuzione di probabilità (meccanismo) della indisponibilità del dato. Ad esempio, il dato manca perché è capitato qualcosa di imprevedibile oppure perchè al di sotto di una soglia o al di sopra di un'altra il dato non può essere osservato? Per poter effettuare inferenze valide senza specificare esplicitamente il meccanismo di mancata risposta, bisogna ipotizzare che questo non dipenda dai valori mancanti: in altri termini si deve supporre che i dati mancanti siano legati al meccanismo che produce i dati completi e che sia assente uno specifico meccanismo che generi i dati incompleti. La presenza di dati mancanti influisce sulle proprietà degli stimatori (media, varianza, percentuali, parametri dei coefficienti di regressione) e pertanto può condurre a risultati inferenziali non corretti. si può pensare al valore mancante come il tassello vuoto in un puzzle. L'aspetto problematico é che siamo a conoscenza che un valore manca, ma ne ignoriamo i contorni esatti: se fosse veramente un puzzle sapremmo che cosa cercare, ma se i



Figura 2.28: Puzzle

contorni rimangono sconosciuti é molto difficile trovare ció che manca. Un' aiuto arriva da conoscenze esterne: il tema del puzzle, ad esempio, oppure da informazioni interne: i colori del contorno. Il completamento delle lacune riscontrate nei dati é simile ad un lavoro di restauro che richiede pazienza, competenze tecniche, ... e ancora pazienza. In questo contesto é determinante stabilire se il meccanismo che ha generato i valori mancanti è di tipo casuale o meno, nonchè analizzare le possibili relazioni tra i valori mancanti e i valori assunti dalle variabili per cui si dispone dell'osservazione.

Definita una matrice R i cui vettori colonna sono delle variabili indicatrici di evento in base alle quali il generico elemento di R, r_{ij}, é pari a 1 se x_{ij} (l'elemento generico di X) è un dato mancante e $r_{ij} = 0$ se x_{ij} é osservato. Tale matrice, nota come matrice delle mancanze (Rubin, 1976), è trattata come un insieme di variabili casuali legate da una distribuzione di probabilità congiunta. Se con X si indica l'intera matrice dei dati, questa può essere partizionata nelle due sottomatrici di dati osservati (X_{obs}) e dati mancanti (X_{miss}): X_{com}=(X_{obs}, X_{miss}). Nella figura che segue è riportato un esempio di un semplice data set affetto da valori mancanti. Più in avanti si riporta la corrispondente matrice delle mancanze. In base alla distribuzione di probabilità della matrice R e alla sua relazione con la matrice dei dati, Rubin (1976) identifica tre tipologie di dati mancanti:

- Missing Completely at Random (MCAR), se la presenza del dato mancante è dovuta al caso e la probabilità di osservare i dati mancanti per un item è indipendente dalle altre risposte date a quello e a gli altri item

$$Pr(R|X_{com}) = Pr(R); \tag{2.20}$$

Gli elementi mancanti sono semplicemente un campione casuale estratto dalla stessa popolazione dei dati validi; è la situazione più comoda e confortevole per la gestione

Record Number	Identification Classification	Weight	Variables				
			V1	V2	V3	V4	V5
1	X	W_1	A	a	y	3.1	4.3
2	X	W_2	A	a	z	4.6	2.8
3	X	W_3	A	b	y	—	1.1
4	X	W_4	B	b	z	2.3	4.6
5	X	W_5	B	c	y	4.9	2.3
6	X	W_6	B	b	—	3.2	3.6
7	X	W_7	C	—	x	3.0	—
8	X	W_8	C	—	y	—	1.2
9	X	W_9	C	a	—	.0	2.4
10	X	W_{10}	—	b	y	—	1.4

Figura 2.29: Esempio di dati mancanti in un data set

Record Number	Identification Classification	Weight	Variables				
			V1	V2	V3	V4	V5
1	X	W1	0	0	0	0	0
2	X	W2	0	0	0	0	0
3	X	W3	0	0	0	1	0
4	X	W4	0	0	0	0	0
5	X	W5	0	0	0	0	0
6	X	W6	0	0	1	0	0
7	X	W7	0	1	0	0	1
8	X	W8	0	1	0	1	0
9	X	W9	0	0	1	0	0
10	X	W10	1	0	0	1	0

Figura 2.30: Esempio di una matrice di mancanze

dei dati mancanti,anche se,purtroppo è quella meno ricorrente.

- Missing at Random (MAR), se la probabilità di osservare un'unità mancante dipende solo dai dati osservati

$$Pr(R|X_{com}) = Pr(R|X_{obs}); \tag{2.21}$$

Il termine MAR è un pò fuorviante perchè le lacune non si verificano proprio a caso, poiché le mancanze potrebbero dipendere da altre variabili del data set. Si tratta di un meccanismo non troppo frequente; ad esempio lo si incontra quando si interrogano persone sul loro stile di vita che rispondono agevolmente a domande sulle loro abitudini alimentari,ma sono poco inclini a rispondere sul loro reddito o su questioni sensibili sulla loro vita privata. Peraltro le risposte possono essere legate ad un'altra variabilequale il grado di istruzione.

- Missing Not At Random (MNAR), se la distribuzione di probabilità associate a R dipende sia dai dati osservati che dalla distribuzione dei dati mancanti

$$Pr(R|X_{com}) = Pr(R|X_{obs}, X_{miss}). \tag{2.22}$$

E' la situazione più frequente, ma anche più difficile da gestire e superare, è tipizzata dalle informazioni sul reddito: la possibilità di avere risposte complete dipende dall'entità del reddito: se è troppo basso o troppo alto c'è una minore disponibilità a dichiararlo. Una situazione simile si configura quando i valori da acquisire sono sotto una soglia di rilevazione o quantificazione (dati censurati).

L'ipotesi che il meccanismo di generazione dei dati mancanti sia MAR implica che i valori mancanti sono prevedibili in base alle risposte osservate nelle altre variabili; al contrario, quando i dati mancanti sono MNAR, l'informazione contenuta nel dataset non è sufficiente per predire ciò che non è stato osservato. In pratica è piuttosto difficile accertare se un particolare data set incompleto sia andato soggeto ad un meccanismo di tip MAR o di tipo MCAR. Diciamo subito che sia l'uno che l'altro richiedono che la variabile lacunosa presenti le stesse caratteristiche sia nei record completi che in quelli incompleti.

La distinzione tra le due strutture passa per il ruolo delle altre variabili nel data set. Se le carenze in una variabile possono essere legate a quelle di un'altra variabile,ma non alla variabile stessa allora il meccanismo è di tipo MAR. Se non c'è legame con le altre variabili il meccanismo è di tipo MCAR. La differenza è inesistente per matrice dei dati con una sola colonna. Il meccanismo MAR può operare solo in data set multivariati; ed infatti le altre variabili possono servire per la ricostruzione dei dati omessi o, per qualche ragione, esclusi. E' perciò importante stabilire quale dei due meccanismi abbia contaminato il nostro data set. sfortunatamente la validazione dell'ipotesi MAR è molto difficile se non impossbile perchè è, di solito, aldilà delle possibilità di controllo del ricercatore. D'altra parte se fosse possibile controllare il flusso dei dati mancanti questi molto probabilmente non si verificherebbero.

Una modalità MAR è meno restrittiva rispetto ad una MCAR dato che la prima consente almeno un legame di dipendenza con altre variabili dove la seconda pretende indipendenza da queste. La MAR è inoltre meno irrealistica, anche se non per questo sempre ammissibile anche se, man mano che aumentano le variabili ed aumenta la loro rilevanza ai fini della spiegazione di quella con valori mancanti, l'ipotesi MAR diventa sempre più ragionevole. Si può affermare che maggiore è la rilevanza e la inter dipendenza dalle altre variabili con la variabile lacunosa più verosimile diventa l'ipotesi MAR.

La figura 2.31 rappresenta graficamente la distinzione appena fatta: se X rappresenta il vettore di variabili completamente osservate, Y una variabile osservata parzialmente (ovvero con dati mancanti), Z il vettore delle cause della presenza dei dati mancanti (non correlate a X e a Y) e R la matrice missingness, si possono più facilmente comprendere le definizioni appena date.

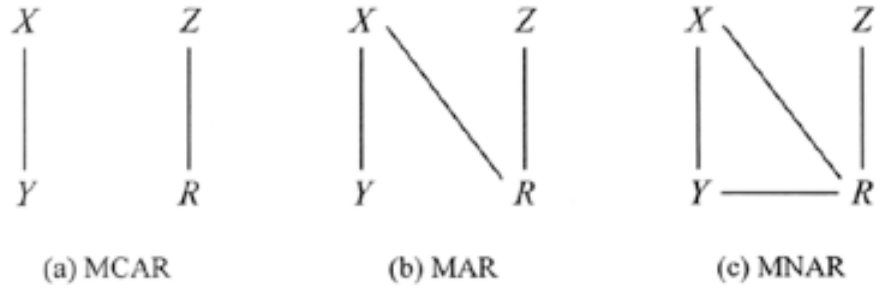


Figura 2.31: Tipologie di dati mancanti

Rappresentazione grafica di (a) missing completely at random (MCAR), (b) missing at random (MAR) e (c) missing not at random (MNAR) (Schafer e Graham, 2002)

2.5.1 Dati mancanti completamente a caso (MCAR)

E' l'ipotesi più facile da trattare, ma è anche quella più raramente riscontrabile in situazioni concrete: la probabilità di osservare una risposta mancante è indipendente sia dalla parte osservata che da quella non osservata dell'insieme di dati completo. Il termine ha un preciso significato (Little & Rubin, 1987; Rubin, 1977): pensando al dataset come un'ampia matrice di dati, i valori mancanti sono casualmente distribuiti attraverso la matrice. Negli studi sulle famiglie questo accade raramente dato che è ben noto che individui in gruppi di minoranze, persone con alti redditi, soggetti con un basso livello d'istruzione e persone soggette a sindromi di depressione o di ansia, sono meno propense a rispondere a tutti gli item di un questionario rispetto alle loro controparti. Anche negli studi clinici è raro osservare dati mancanti completamente a caso, perchè gli addetti all'inserimento dei dati solitamente trascurano alcuni campi in modo sistematico, o perchè i pazienti con particolari condizioni piuttosto che altre si rifiutano di fornire particolari informazioni.
Nella figura 2.32 Y_N indica la variabile *target* in una popolazione di N unità; inoltre $\overline{Y}_n$: media campionaria (stimatore della media nella popolazione) su un campione casuale semplice di n unità; $\overline{Y}_n = \dfrac{(\sum_{i=1}^{n} y_i)}{n}$
Se ci sono solo r rispondenti lo stimatore diventa $\overline{Y}_r$ calcolato su r unità anziché n. Se i rispondenti possono essere considerati un campione casuale semplice delle unità, lo stimatore $\overline{Y}_r$ è corretto. C'è solo una perdita di precisione dello stimatore. In effetti la varianza aumenterebbe di un fattore $\approx n/r$. In questo caso si dice che il meccanismo di mancata risposta è completamente casuale (MCAR, Missing Completely at Random).

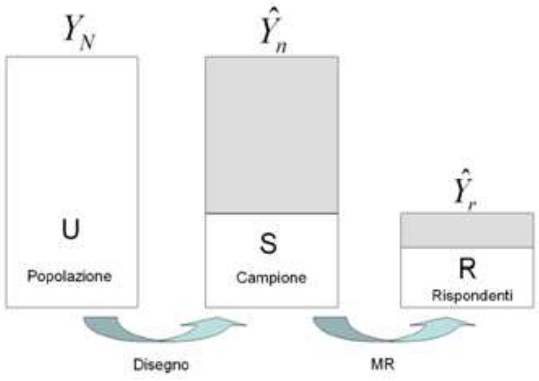


Figura 2.32: MCAR

2.5.2 Dati mancanti a caso (missing at random, MAR)

Si parla di dati MAR o non risposta ignorabile quando la probabilità di osservare una risposta mancante dipende soltanto dalla parte osservata dell'insieme di dati. I dati mancanti su una variabile specifica si definiscono "mancanti a caso" se la verosimiglianza dei dati mancanti sulla variabile d'interesse non è correlata con il punteggio individuale su quella variabile, dopo aver controllato le altre variabili nello studio. In uno studio sulla depressione materna (Acock, 2005), il 10% o più delle madri può rifiutare di rispondere alle domande sul loro livello di depressione.

Si supponga che uno studio includa lo stato di povertà codificato come 1="in stato di povertà" e 0="non in stato di povertà". Un punteggio sulla depressione della madre è "mancante a caso" se i suoi valori mancanti sulla depressione non dipendono dalla variabile "livello di depressione", tenendo sotto controllo la variabile "stato di povertà". Se la verosimiglianza del rifiuto a rispondere al quesito sul livello di depressione è correlato con lo stato di povertà, ma non è correlato con il livello di depressione entro ogni classe di stato di povertà, allora i valori mancanti sono detti di tipo MAR. Per i dati MAR, in questo esempio, il punto focale non è se lo stato di povertà possa predire la depressione materna, ma se lo stato di povertà rappresenti un "elemento esplicativo" della presenza o meno dell'informazione "livello di depressione materna". Una variabile è considerata "elemento esplicativo" quando aiuta a spiegare se un soggetto risponderà o meno ad un quesito (Raghunathan, 2004; Schafer, 1997). Molti "elementi esplicativi" vengono inclusi negli studi sulle famiglie di grande scala: i più comuni "elementi esplicativi"inclusi sono il livello d'istruzione, la razza, l'età, il sesso ed indicatori di benessere psico-sociale. L'assunto per i valori MAR è valido solo se il modello dei dati mancanti é condizionatamente casuale, dati

i valori osservati nelle variabili considerate "elementi esplicativi".

La tabella riportata in figura 2.33 descrive la situazione in cui si riscontrano dati mancanti su due variabili: età e peso.

Record	Genere	Età	Peso	Re	Rp
1	0	28		0	1
2	1	19	218	0	0
3	1	37	235	0	0
4	0	24	150	0	0
5	1	18		0	1
6	1		176	1	0
7	1			1	1
8	0			1	1
9	0	28		0	1
10	1	46	195	0	0
11	0	23		0	1
12	0	29		0	1
13	1	44	221	0	0
14	0			1	1
15	0	21		0	1
16	1	41	204	0	0
17	0	40		0	1
18	1	37	208	0	0
19	0			1	1
20	1	43		0	1

Figura 2.33: MAR per due variabili

Se si riscontra un grado di associazione elevato rispettivamente tra R_e e genere e/o tra R_p e genere allora il meccanismo di non reporting si può considerare di tipo MAR. Nell'esempio la statistica χ^2 è significativa per la seconda associazione (valor $p = 0.002$), ma non per la prima (valor $p = 0.026$);quindi c'è sia un MAR che un MCAR.

2.5.3 Dati mancanti non a caso (MNAR)

Si parla di dati MNAR o non ignorabili quando la probabilità di risposta dipende sia dai dati osservati che da quelli non osservati: in questo caso il meccanismo della non risposta deve essere tenuto esplicitamente in considerazione nel modello di analisi del fenomeno. Per esempio se sis sta analizzando la salute mentale di un gruppo di persone a cui è stata diagnosticata la depressione bisogna considerare che le persone ammalate tendono a non dichiarare il loro stato mentale. E' evidente che queste lacune non avvengono a caso e che un eventuale punteggio medio sullo status mentale calcolato solo sui dati disponibili fornirà una stima distorta della media che si sarebbe ottenuto disponendo di tutti i dati.

In uno studio sugli iscritti all'università in cui sia centrale la performance accademica si deve tenere conto vhe coloro che si ritirano e generano valori mancanti hanno

ottenuti risultati modesti negli esami a cui si sono presentati. Quindi la modalità di formazione dei vuoti nei dati è di tipo MNAR. Analogamente, nel riportare la percorrenza di un veicolo si è forse portati a non fornire distanze che denotino una vetustà giudicata eccessiva e la dinamica deve essere considerata MNAR anche in questo caso. Una situazione più complessa è quella che si configura nel rispondere a domande ordinate secondo la difficoltà di risposta. Se gli ultimi *item* non ricevono risposta si possono considerare due casi: non risposta perché chi risponde non è in grado di darla o non risposta perché il soggetto sceglie di non darla. Nel primo caso il meccanismo è di tipo MNAR, nel secondo caso è di tipo MCAR.

La modalità MNAR non consente di ignorare la dipendenza dei valori mancanti dalla variabile in cui si riscontrano: significherebbe trascurare informazioni fondamentali con un atteggiamento poco scientifico. Ignorare che la tendenza a riportare l'entitè dei difetti di produzione varia in ragione di tale entitè comporebbe formulazioni di modelli sbagliati o inadeguati per spiegare il loro impatto. Negli studi longitudinali in cui una coorte di persone è seguita per un periodo prolungato di tempo, si è accertato che, almeno negli USA, i soggetti sposati, anziani, bianchi, istruiti e le donne tendono a permanere in modo più duraturo nello studio rispetto ad altre categorie di soggetti.

Quando si verificano situazioni di questo tipo l'applicazione delle tecniche di imputazione diventa assai più problematica. Ad esempio un partecipante ad uno studio sulla perdita di peso non si presenta ad una delle verifiche periodiche, i dati relativi potrebbero mancare per fattori non trascurabili: in contrasto con il caso precedente, il meccanismo che determina la perdita di informazioni può essere spiegato soltanto dalla variabile che presenta mancanze. La locuzione "non ignorabile" significa che anche tenuto conto di tutte le informazioni osservate disponibili la ragione perchè alcuni dei dati manchino dipendono dai dati che mancano. Pertanto, se il meccanismo è di tipo MNAR la gestione dei dati mancanti è molto difficoltosa.

Il solo modo di ottenere stime non distorte dei parametri di interesse o comunque per ricostruire in modo credibile i dati non riportati è necessario predisporre di un buon modello che spieghi gli aspetti "not random" del vuoto nei dati. Quando questo modello è disponibile lo si può incorporare in una procedura complessa per stimare i valori mancanti. Sfortunatamente nello scenario MNAR è raro che si possa modellare correttamente l'influenza sistematica che subisce il data set.

In un contesto MNAR la causa dei dati mancanti potrebbe essere legata a qualche aspetto sconosciutoo non osservabile che può complicare notevolmente l'analisi. Nella figura che segue sono considerate due variabili relative a rilevazioni atmosferiche in una zona tropicale: temperatura alla superficie del mare e umidità. Nella

parte sinistra della figura sono riportate le combinazioni della matrice delle mancanze. Nel secondo e quarto quadrante si vedono le assenze singole, nel terzo quadrante si vedono le coassenze.

Nel grafico a destra si può notare che le mancanze singole sulla temperatura sono maggiormente associate con valori elevati del vento in quota (componente uwind).

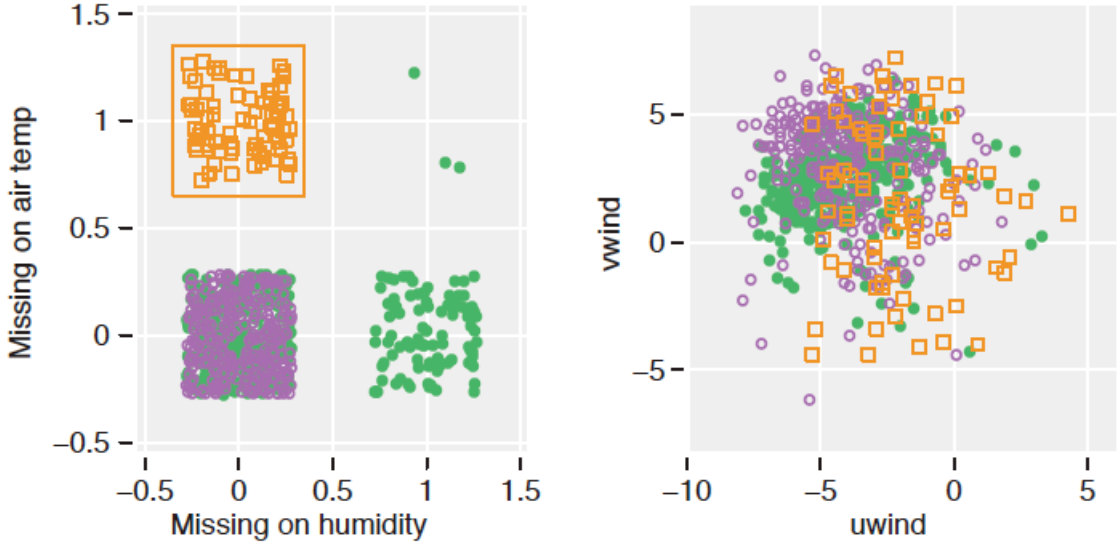


Figura 2.34: Esempio di meccanismo MNAR informativo

I valori che mancano non sono assenti per caso. La loro trattazione necessita lo studio della dipendenza tra missing e non missing.

2.6 Valori mancanti in una serie storica

L'analisi delle serie storiche offre il vantaggio di individuare le tendenze in atto in un fenomeno monitorato nel tempo. Quindi i metodi che hanno come oggetto le serie storiche debbono essere improntati a questa caratteristica dei dati. Solo in parte possiamo adoperare i concetti dei valori mancanti discussi in precedenza. La situazione tipica nelle serie storiche è quella descritta nella figura 2.35
I metodi statistici per le serie storiche presuppongono che si disponga di una informazione valida su tutte le unità di tempo incluse nell'analisi. Quindi, i dati mancanti riducono la portata dell'indagine, cioè la profondità e l'estensione dell'inferenza basata sul *data set*. Può essere completa la serie storica che precede il dato mancate e completa la serie storica che segue, ma non tutta la serie storica. Ad esempio, le autocorrelazioni semplici e parziali descritte nel primo capitolo possono essere calcolate con la clausola di limitare la formula alle sole coppie di valori effettivamente presenti

```
data usprice;
   input date monyy. cpi ppi
   format date monyy.;
datalines;
jun1990     . 114.3
jul1990     . 114.5
aug1990 131.6 116.5
sep1990 132.7 118.4
oct1990 133.5 120.8
nov1990     .     .
dec1990 133.8 118.7
jan1991 134.6 119.0
feb1991 134.8 117.2
mar1991 135.0     .
apr1991 135.2 116.0
may1991 135.6 116.5
jun1991 136.0 116.3
jul1991 136.2     .
```

Figura 2.35: Esempio di serie storica con valori mancanti

in entrambe le componenti. Così facendo però si producono due inconvenienti: si riduce la attendibilità dei valori che decade in ragione del numero di coppie utilizzate e si ignorano anche dati validi (una distorsione o *bias* che si aggiunge a quello della riduzione fisiologica del numero di coppie dovuta all'aumentare del ritardo). Senza poi contare quello che succede quando si procede al calcolo delle differenze prime, seconde, etc. che diventano impraticabili in caso di uno o più dati mancanti.

I valori mancanti sono informazioni non reperibili, omesse ovvero negate oppure perse per una unità statistica della quale però sono disponibili altre informazioni valide (quali i dati delle unità temporalmente contigue). I valori mancanti, qualunque ne sia la ragione preoccupano perchè si riduce il numero di casi validi il che, a sua volta, è considerata una iattura per moltissime analisi statistiche (ad esempio le stime campionarie e la verifica di ipotesi); inoltre la indisponibilità di certe informazioni e non di altre potrebbe creare serissime distorsioni allorchè ci siano elementi per ritenere che i dati non riportati siano non stocasticamente diversi da quelli riportati.

La soluzione necessaria al problema dei dati mancanti nei dati temporali è di riempire il vuoto completando la serie storica in modo da renderla disponibile per analisi successive (procedura di imputazione). Occorre però fare subito alcune precisazioni. La ricostruzione dei valori mancanti è una tecnica inferenziale: a partire da una frazione dei casi osservati e potenzialmente osservabili, si tenta di ottenere informazioni relative alla totalità dei valori e poi da queste ridiscendere verso i valori mancanti per ricostruirli. Ad esempio, si calcola la media aritmetica dei valori effettivamente osservati della serie storica e la si colloca al posto dei valori mancanti. Il fatto che alcune delle tecniche proposte abbiano una natura deterministica e non

stocastica non è rilevante per togliere all'imputazione la natura inferenziale.

L'imputazione introduce dati artificiali e si deve garantire che, attraverso di essa, non siano state introdotte artificiosità negative. Questa garanzia può essere data in relazione alla finalità con cui i dati sono stati richiesti. Attenzione! Qualunque sia il meccanismo di imputazione del dato mancante a partire dai soli dati completi, il presupposto indiscutibile è che i dati completi siano corretti. Un vecchio brocardo della statistica avverte che non esistono procedure puramente matematiche, per quanto sofisticate, che possano trasformare un dato "cattivo",ma insospettabile in un dato "buono".

L'analisi delle serie storiche presenta delle differenze rispetto alle situazioni tipiche in cui si affrontano i problemi di imputazione. Infatti, l'obiettivo non è di stimare dei parametri di sintesi, ma di proporre dei valori sostitutivi ad altri non osservati che ben si collocano nel flusso della serie storica. In altre parole, l'esito dell'imputazione non si esaurisce nel calcolo di indicatori riassuntivi, ma i valori imputati rimangono nella serie storica che, una volta completata, diventa disponibile per ogni altro interesse cognitivo.

2.6.1 Imputazione dei dati mancanti

Sebbene tutti i metodi di imputazione si basino implicitamente o esplicitamente sull'ipotesi che le mancate risposte siano di tipo MCAR oppure MAR, spesso può capitare di aver a che fare con dati che non appartengono a tali categorie. Un modo per far fronte a questa situazione è quello di lavorare su sottoinsiemi omogenei dei dati che non presentino tali caratteristiche. Nel nostro caso, trattiamo valori mancanti che non abbiano ragioni sistematiche di essere assenti. Ipotesi: MCAR (*missing completely at random*). Le posizioni temporali in cui i dati sono stati omessi formano un campione casuale semplice delle posizioni della serie storica. L'ipotesi è comoda dal punto di vista operativo, ma è poco realistica. C'è sempre una buona ragione per cui un dato non è stato rilevato, anche se non sempre si riesce a conoscerla. Tuttavia, le tecniche di analisi delle serie storiche sono più facili da usare se la serie è completa in tutte le sue osservazioni. I dati mancanti debbono essere ricostruiti perché le discontinuità bloccano molte analisi.

L'idea dell'imputazione è allo stesso tempo fascinosa e fuorviante. Seduce perché da l'impressione, o meglio lascia intendere, che si abbia di fronte un profilo di serie storica completo, il che è solo una credibile illusione, non una realtà. E qui c'è il rischio di essere ingannati dall'uso di un serie che è completa solo in apparenza. Tuttavia non si deve trascurare né il vantaggio né la comodità di poter operare su un

blocco di dati completo, soprattutto quando si intendono applicare tecniche, quali la metodologia Box-Jenkins, che non prevedono dati mancanti. Si deve aggiungere che una parte consistente delle imputazioni è realizzata da persone o organizzazioni molto vicine alla raccolta dei dati. Certo molto più vicine degli utenti finali che possono disporre solo di una parte limitata delle serie storiche di interesse.

Supponiamo che la serie storica W_t sia priva del dato relativo alla posizione t-esima. La fase cruciale dell'indagine è stata la scelta del metodo di individuazione del valore più ragionevole da attribuire a W_i. Le tradizionali procedure di imputazione dei dati mancanti hanno come base di riferimento l'intera serie storica. Ad esempio, la sostituzione con la media aritmetica o con la mediana si realizza calcolando la media aritmetica su tutti i valori completi; il valore così ottenuto va a collocarsi al posto di tutti i dati mancanti. Si vedano le figure 2.36 e 2.37

Anno	Mese	Yt	
2006	jan	95	95
2006	feb	95	95
2006	mar	94	94
2006	apr	106	106
2006	may	97	97
2006	jun	103	103
2006	jul	**na**	**101**
2006	aug	95	95
2006	sep	102	102
2006	oct	105	105
2006	nov	97	97
2006	dec	100	100
2007	jan	99	99
2007	feb	96	96
2007	mar	103	103
2007	apr	96	96
2007	may	102	102
2007	jun	108	108
2007	jul	102	100
2007	aug	**na**	**101**
2007	sep	91	91
2007	oct	107	107
2007	nov	107	107
2007	dec	104	104
2008	jan	103	103
2008	feb	104	104
2008	mar	109	109
2008	apr	97	97
2008	may	103	103
2008	jun	105	105
2008	jul	93	93
2008	aug	103	100
2008	sep	95	95
2008	oct	99	99
2008	nov	108	108
2008	dec	102	102
2009	jan	94	94
2009	feb	101	101
2009	mar	104	104
2009	apr	109	109
	Media	**101**	
	Mediana	**102**	
	SQ	**100.13**	

Figura 2.36: Imputazione con il metodo della media

Le conseguenze di questa procedura sono evidenti: sottostima della variabilità, dal momento che si aggiungono valori tutti uguali; interruzione dei fattori di trend, stagionalità ciclo, perché si aggiunge un valore artificialmente stabile; aumento della curtosi, data la ripetizione dei valori al centro; minore attendibilità dell'inferenza statistica a causa di questi problemi che riducono la gaussianità dei dati.

La media potata è un antica procedura di media robusta che prevede il calcolo di quest'ultima dopo aver escluso il donatore, o un gruppo di donatori, con il valore più piccolo e il donatore con il valore più grande. È un sicuro miglioramento rispetto alla

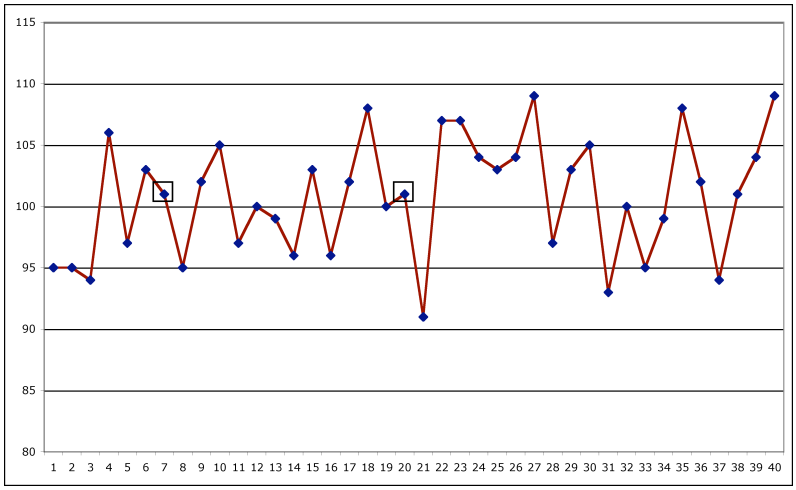


Figura 2.37: Serie storica con valori imputati

media globale, ma non se ne conosce l'impatto effettivo. Tendenzialmente si taglia in ambedue le direzioni, in gran parte ciò è dovuto ad un ben noto risultato degli intervalli di confidenza: intervalli simmetrici hanno un diametro inferiore a parità di livello di confidenza. Naturalmente questo è vero se e solo se la distribuzione dei dati è simmetrica. In mancanza di tale presupposto i tagli asimmetrici potrebbero essere preferibili. D'altra parte, né gli intervalli simmetrici né quelli asimmetrici possono proteggerci dalle anomalie che provengono della distribuzione (valori grandi o valori piccoli che siano) da un solo lato e inoltre l'ipotesi implicita e non sempre condivisibile delle medie potate è che esista una frazione fissa di valori che necessariamente sia da scartare.

$$\hat{W}_t = \frac{\sum_{i=\lfloor \gamma_1 n \rfloor + 1}^{\lfloor \gamma_2 n \rfloor - 1} W_{(i)}}{n - \lfloor \gamma_1 n \rfloor - \lfloor 1 - \gamma_2 n \rfloor - 2} \qquad 0 < \gamma_1 < \gamma_2 < 1 \tag{2.23}$$

Nel caso della media potata agli estremi $\gamma_1 \times n = 1$ e $\gamma_2 \times n = n$. In questo caso si prevede il calcolo della media dopo aver escluso il donatore con il valore più piccolo e il donatore con il valore più grande. Quando $\gamma_1 = 0.25$ e $\gamma_2 = 0.75$ si ottiene la media potata al primo e al terzo quartile, mentre se $\gamma_1 = 0.1$ e $\gamma_2 = 0.9$ otteniamo la media potata dall'esclusione del primo decile e dell'ultimo decile. La mediana può essere vista come un caso estremo della media potata, quando solo solo uno o due valori centrali sono mantenuti.

Una possibilità da considerare è la mediana Hodges e Lehmann

$$\hat{W}_t = Me \left\{ \frac{W_i + W_j}{2}; \qquad i < j \right\} \tag{2.24}$$

dove i valori nella variabile *target* sono disposti in ordine. Questo indicatore può essere visto come l'unificazione delle caratteristiche positive di media aritmetica e

mediana in cui, come al solito, il calcolo è esteso ai soli valori validi. Il valore così ottenuto è sostituito a tutti i valori mancati della serie storica indipendentemente dalla posizione occupata e dal loro vicinato (sono le debolezze del metodo).

Esempio 5. *Consideriamo una serie storica breve relativa alle vendite trimestrali del manuale dell'SPSS.*

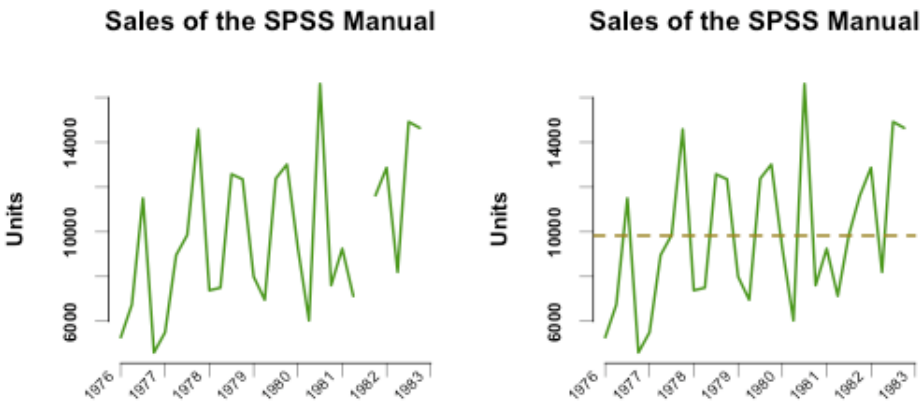


Figura 2.38: Imputazione con la mediana di Hodges-Lehmann

Il calcolo della mediana di Hodges e Lehmann porterebbe a imputare 9857 *invece di* 15051 *con un errore del* 35% *che forse è troppo elevato per accontentarsi.*

Tutte le tecniche illustrate in precedenza ignorano l'ordinamento temporale dei valori e la loro eventuale articolazione in stagioni e substagioni. Una procedura di imputazione molto semplice è il metodo della tabella: i dati stagionali si organizzano in una tabella 2 con righe/anno e colonne/mesi.

Si può imputare sia la media della "stagione" che la media dell'anno ovvero la media tra quella annuale e quella stagionale. Analogo ragionamento si può fare usando la mediana. Il metodo dà poche garanzie dato il ragionamento molto elementare alla sua base. E' solo un po' meglio della imputazione della media o della mediana globale. Esempi di queste applicazioni sono contenuti nella figura 2.39.

L'interpolazione lineare tra due valori contigui è pure una procedura grezza, ma che ha almeno il merito di tenere conto della collocazione temporale del valore omesso, almeno rispetto ai casi contigui.

$$\hat{W}_t = W_{t-1} + \frac{W_{t+1} - W_{t-1}}{2} \tag{2.25}$$

L'ipotesi è che entrambi i valori contigui siano disponibili.

Anno	jan	feb	mar	apr	may	jun	jul	aug	sep	oct	nov	dec
2006	95	95	94	106	97	103	**98**	95	102	105	97	100
2007	99	96	103	96	102	108	102	**99**	91	107	107	104
2008	103	104	109	97	103	105	93	103	95	99	108	102
2009	94	101	104	109								

Anno	jan	feb	mar	apr	may	jun	jul	aug	sep	oct	nov	dec
2006	95	95	94	106	97	103	**99**	95	102	105	97	100
2007	99	96	103	96	102	108	102	**102**	107	107	107	104
2008	103	104	109	97	103	105	93	103	95	99	108	102
2009	94	101	104	109								

Anno	jan	feb	mar	apr	may	jun	jul	aug	sep	oct	nov	dec
2006	95	95	94	106	97	103	**99**	95	102	105	97	100
2007	99	96	103	96	102	108	102	**101**	107	107	107	104
2008	103	104	109	97	103	105	93	103	95	99	108	102
2009	94	101	104	109								

Figura 2.39: Esempio del metodo di imputazione tabellare

Anno	jan	feb	mar	apr	may	jun	jul	aug	sep	oct	nov	dec
2006	95	95	94	106	97	103	**97**	95	102	105	97	100
2007	99	96	103	96	102	108	102	**105**	107	107	107	104
2008	103	104	109	97	103	105	93	103	95	99	108	102
2009	94	101	104	109								

$$Y_{jul,2006} = 103 + \frac{(95-103)}{2} = 103 - \frac{8}{2} = 103 - 4 = 97$$

$$Y_{aug,2007} = 102 + \frac{(107-102)}{2} = 102 + \frac{5}{2} = 102 + 2.5 = 105$$

Figura 2.40: Esempio del metodo di imputazione con interpolazione

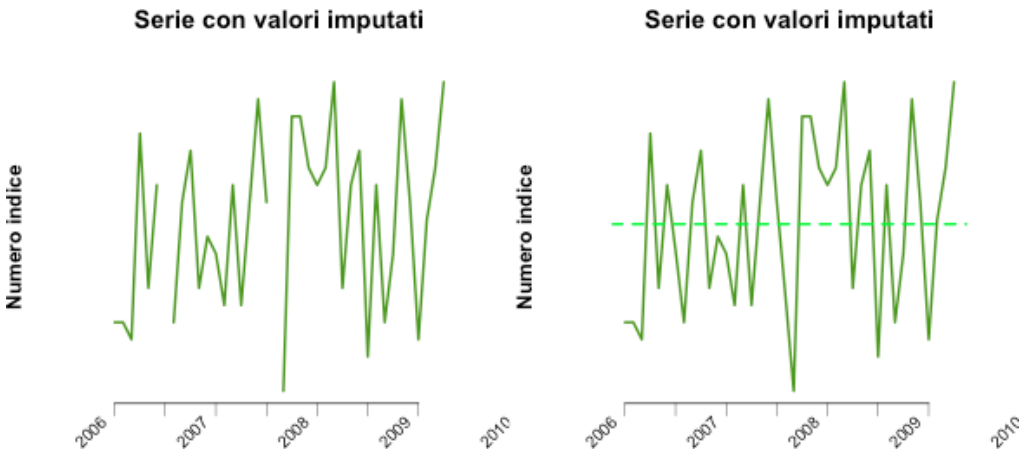


Figura 2.41: Esempio del metodo di imputazione stagionale

La procedura implementata nel nostro *software* tiene conto della stagionalità -se presente- ed infatti trova il valore valido stagionale precedente il dato mancante e quello stagionale susseguente come media aritmetica semplice di tutte le osservazio-

nii stagionali valide nella serie storica. Per i dati non stagionali calcola la media aritmetica potata del 25% inferiore e superiore sia per i valori posizionati indietro rispetto al mancante che per quelli posizionati in avanti. La potatura è necessaria per sterilizzare le conseguenze negative dei valori anomali. Il dato è poi imputato con la media ponderata di questi valori

$$\hat{W}_t = \frac{W_s^- + W_s^+ + 4\left(W_t^- + W_t^+\right)}{10} \tag{2.26}$$

Per $t = 1$ si pone $W_t^- = W_2$ e per $t = n$ si pone $W_t^+ = W_{n-1}$. La sequenzialità delle osservazioni è considerata grazie alle medie delle osservazioni stagionali; anche le medie globali (pre e post) valore mancante sono in linea con il carattere evolutivo dei dati. Tuttavia, la *naiveté* della procedura salta subito agli occhi e non manca di destare perplessità. A difesa possiamo dire che, l'imputazione è meglio della cancellazione del periodo con relativo trattamento di una serie storica più breve e irregolare rispetto alle unità di tempo (questa soluzione non va bene per serie stagionali). Se poi la scelta dei valori imputati è insoddisfacente, i valori possono essere stimati con un primo modello e poi ancora ristimati come se le stime fossere dei valori osservati. Non a caso, tra i metodi per trattare i valori mancanti nelle serie storiche c'è la metotologia Box-Jenkins che però dovrebbe essere applicata per interpolare i valori ed essere ri-usata per asseverarne il risultato.

Esempio 6. *Consideriamo una serie storica breve relativa all'ambiente R.*

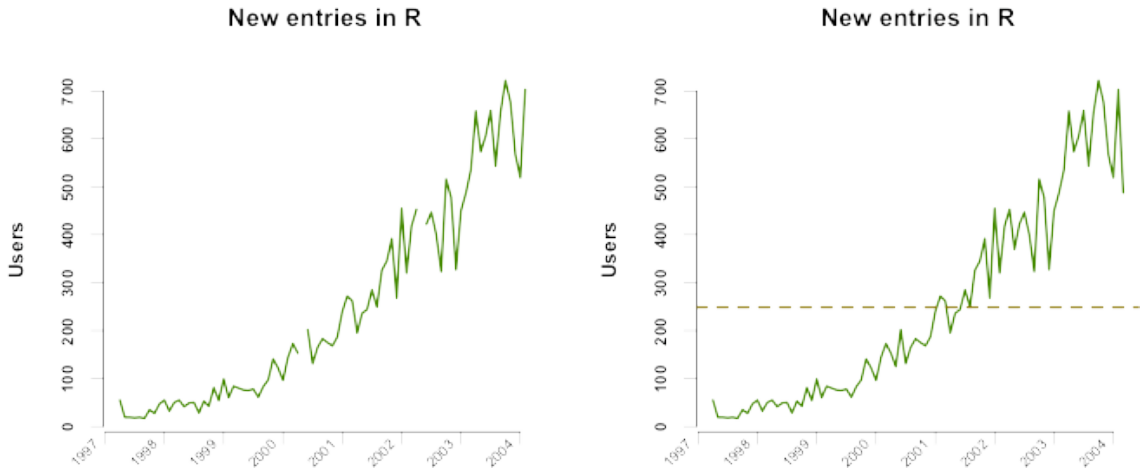


Figura 2.42: Imputazione "home made

Il nostro metodo imputa $124.7, 369.5, 487.5$. *I valori veri sono* $226, 468, 886$. *Gli errori sono elevati, intorno al* 50%, *ma se non altro non si usano i dati per costruire un modello che genera i dati su cui si costruisce un'altro modello o lo stesso.*

E' sempre necessario intervenire per la riparazione della serie storica: valori troppo diversi dall'ordinario possono avere effetti imprevedibili e duraturi; valori mancanti possono oscurare la percezione di movimenti importanti. L'aggiustamento della serie storica avviene con valori sostitutivi o di rimpiazzo di quelli "sbagliati". Il presupposto con cui sono scelti è che vuoti ed anomalie non siano affioramenti di comportamenti aberranti del fenomeno che si studia. In caso contrario possono risultare pezze a colori che coprono bene la rottura, ma nascondono il vero problema.

2.7 Differenziazione di una serie storica

L'analisi delle serie storiche ci pone spesso di fronte a realizzazioni di processi soggetti a fattori cumulativi, ovvero forze -anche esogene- che spingono il processo a seguire particolari strutture di crescita: andamento demografico, inflazione, progresso tecnico, avanzamenti culturali, etc. Ad esempio una serie monetaria misurata a prezzi correnti esibirà un trend crescente (probabilmente lineare o lineare nei logaritmi) dovuto all'inflazione. Ed ancora, l'andamento secolare della concentrazione dei redditi sembra mostrare un trend quadratico (a forma di U rovesciata).

La presenza del *ciclo-trend*, componenti importanti di una serie storica, si concreta nel fatto che le variabili casuali allineate nel processo hanno una media non più costante, ma che cambia a seconda della posizione della variabile. Questo implica la non stazionarietà in media del processo che è invece uno dei requisiti minimi richiesti per poter effetuare l'analisi Box-Jenkins.

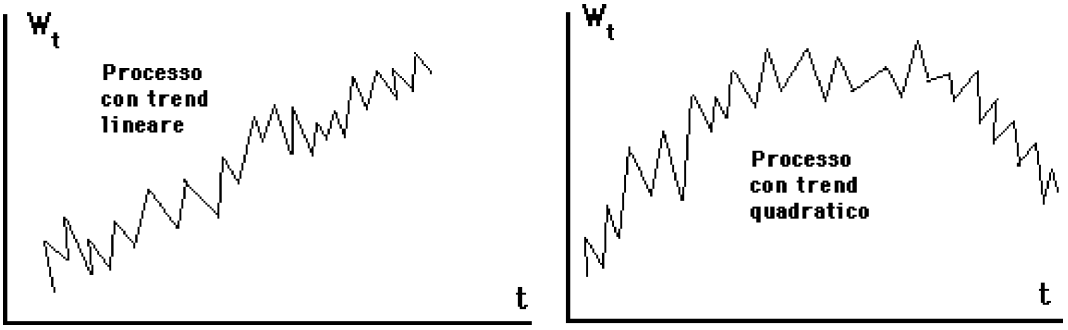


Figura 2.43: Esempio di trend quadratico

Un modo semplice di rimuovere un ciclo-trend polinomiale per indurre la stazionarietà in media è di trasformare la serie originale in differenze prime, seconde, etc. A questo fine è utile introdurre l'operatore di ritardo o di *lag* indicato con B che assume un ruolo importante per il trattamento algebrico delle serie temporali.

Si definisce attraverso la trasformazione

$$BW_t = W_{t-1} \tag{2.27}$$

Se lo stesso operatore è applicato a W_{t-1} produce

$$BW_{t-1} = B\left(BW_t\right) = B^2 W_t = W_{t-2} \tag{2.28}$$

e così via. In generale $B^j W_t = W_{t-j}$. Per completare logicamente la definizione dell'operatore di *lag* si conviene che $B^0 Z W_t = W_t$.

L'operatore differenza Δ è definito attraverso la relazione

$$\Delta = (1 - B) \tag{2.29}$$

Tale operatore trasforma la serie W_t nella serie degli scarti $(W_t - W_{t-1})$. Infatti

$$\Delta W_t = (1 - B)\, W_t = W_t - BW_t = W_t - W_{t-1} \tag{2.30}$$

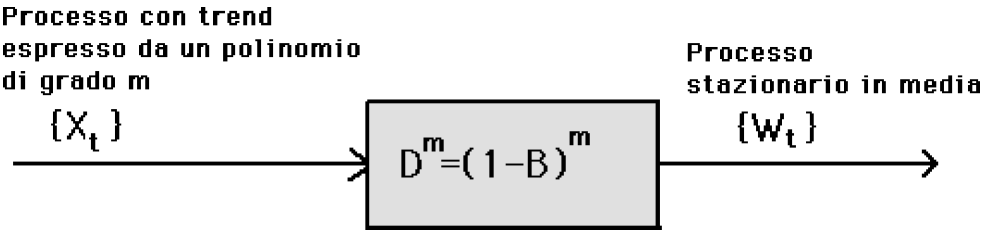


Figura 2.44: Effetto del filtro delle differenze

Allo stesso modo Δ^2 trasformerà W_t nella serie delle differenze seconde:

$$\begin{aligned} \Delta^2 W_t = (1 - B)^2 W_t &= \left(1 - 2B + B^2\right) W_t = W_t - 2W_{t-1} + W_{t-2} \\ &= (W_t - W_{t-1}) - (W_{t-1} - W_{t-2}) \\ &= \Delta W_t - \Delta W_{t-1} \end{aligned} \tag{2.31}$$

Generalizzando l'operatore alla differenza di ordine d avremo

$$\begin{aligned} \Delta^d W_t &= (1 - B)^d W_t = \sum_{i=0}^{d} \binom{d}{i} (-1)^{d-i} B^i \\ &= \left[1 - dB + \binom{d}{2} + \cdots + \binom{d}{i} (-1) B^i + \cdots + (-1) B^i\right] W_t \end{aligned} \tag{2.32}$$

Se la serie fosse costituita dal solo polinomio di grado d, le differenze di ordine d sarebbero tutte costanti. Questo però è valido solo nel modello additivo. In quello

moltiplicativo si procede allo stesso modo, trasformando preliminarmente la serie in scala logaritmica. Inoltre, è facile vedere che se il valore assoluto di B, cioè $|B|$ è inferiore all'unità, per la proprietà della serie geometrica, avremo

$$\Delta^{-1} = \frac{1}{(1-B)} = B^0 + B^1 + B^2 + \cdots + \tag{2.33}$$

per cui $\Delta^{-1}W_t = W_t + W_{t-1} + W_{t-2} + \cdots +$ ovvero l'inversa dell'operazione di differenza è la somma di infiniti termini. Uno dei meriti del fondamentale testo di Box e Jenkins è quello di avere semplificato il modo in cui si rende stazionaria in media una serie fortemente evolutiva: applicare più volte l'operatore differenza.

Per decidere quante volte differenziare una serie si parte dal fatto che

$$\Delta^d \left[\beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_d t^d\right] = costante \tag{2.34}$$

Se quindi le differenze di ordine d oscillano senza struttura intorno ad una costante allora il grado del polinomio dovrebbe essere d. Poiché la differenza di ordine d elimina un trend polinomiale di grado d la serie delle differenze seconde dovrebbe avere minore variabilità della serie delle differenze prime, ma maggiore della serie delle differenze terze. Tuttavia, tale indicazione non sempre è corretta.

Esempio 7. *Consumo residenziale di acqua a Londra. Dati quadrimestrali rilevati a valle di una tariffazione che incoraggia il risparmio.*

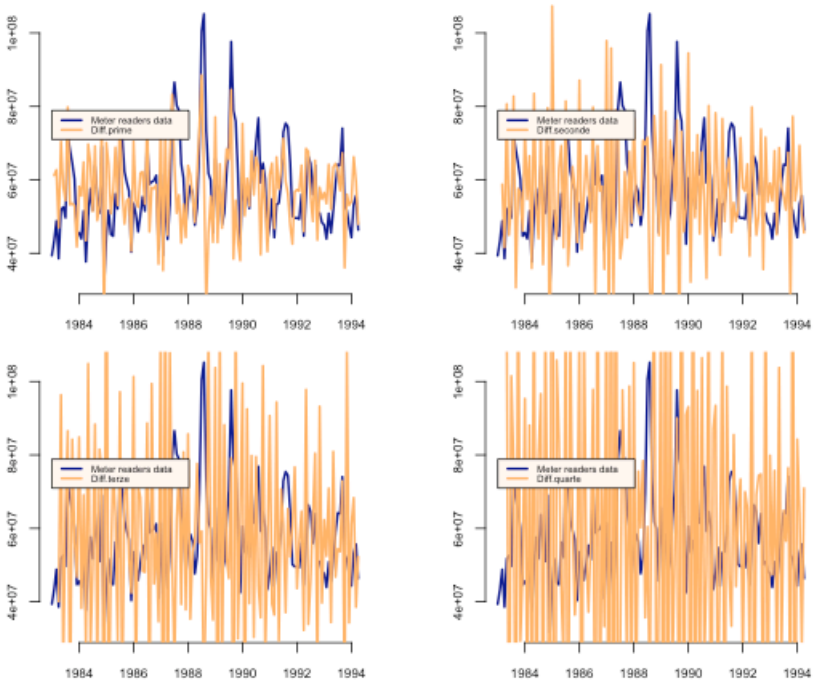


Figura 2.45: Differenziazione regolare della serie storica

Qui è già in dubbio la differenza prima perché non sembra modificare l'andamento della serie originale. Non è il caso di provare differenziazioni maggiori, anche se forse un trend quadratico è presente. In questa fase è importante ricordare che molta della teoria dei processi stocastici è rivolta ai processi stazionari e, pertanto, occorre intervenire in tutti i modi appropriati per ottenere la condizione di stazionarietà.

Esempio 8. *Il trend della serie storica è quadratico.*

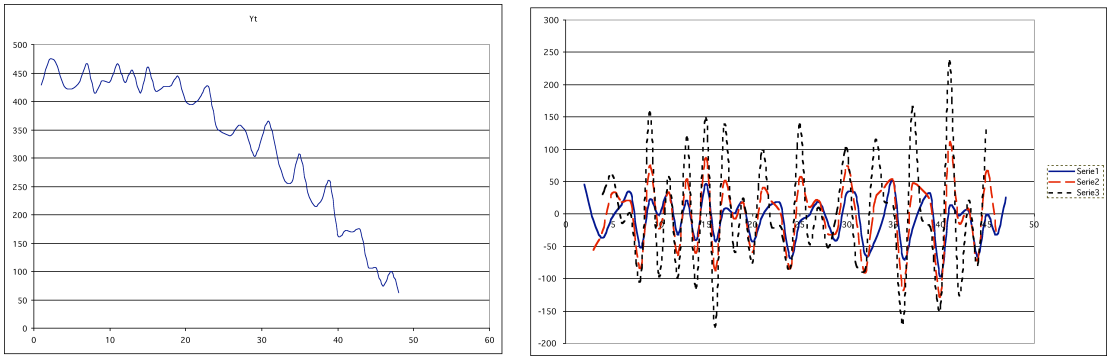


Figura 2.46: Esempio di eliminazione di un trend quadratico

Le differenze prime sembrano già stabili, ma sono quasi tutte negative. Le differenze terze oscillano paurosamente superando in ampiezza addirittura il livello medio della serie storica. Le differenze seconde sono più rispondenti ai canoni di stabilità, ma neanche loro convincono troppo.

Esistono anche dei test per una decisione inferenziale su d, ma l'analisi grafica è di solito sufficiente per valutare correttamente. In caso di dubbio sarà bene scegliere il grado minore perché una differenziazione eccessiva può introdurre delle "periodicità" nella serie storica che potrebbero essere confuse con delle componenti strutturali. Un suggerimento, da considerare con sano scetticismo, è di continuare a differenziare la serie fino a quando la variabilità dei valori diminuisce. Tenuto conto che se la serie presenta auto-correlazioni elevate (in valore assoluto) per *lag* elevati, allora l'ordine delle differenze potrebbe essere altrettanto elevato.

CAPITOLO 3

Processi stocastici

3.1 Introduzione

Il presupposto dell'analisi delle serie storiche è che i valori osservati di un certo fenomeno conservino "memoria" dei fatti antecedenti, siano cioè un effetto cumulativo della loro storia passata a meno di rotture insanabili o di cesure storiche ed eventi destrutturanti. Ci si aspetta quindi che osservazioni molto vicine siano dipendenti o almeno correlate ed osservazioni a *lag* distanti siano prive di legami, o che almeno non abbiano più legami di tipo lineare. La conservazione degli effetti porta a pensare che il fenomeno possa essere governato da un meccanismo in parte deterministico e quindi conoscibile ed in parte stocastico e forse modellabile con la teoria della probabilità ai fini della previsione, del controllo e dell'analisi di scenario del fenomeno descritto dalla serie storica.

Per descrivere analiticamente i fenomeni con memoria si può usare il concetto di processo stocastico lineare che vede la variabile di ogni periodo come una combinazione lineare del valore presente e dei valori passati di un processo *white noise*:

$$\begin{aligned} W_t &= a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \psi_3 a_{t-3} + \cdots \\ &= a_t + \psi_1 B a_t + \psi_2 B^2 a_t + \psi_3 B^3 a_t + \cdots \\ &= \left(1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \cdots\right) a_t \end{aligned} \tag{3.1}$$

dove B è l'operatore di ritardo discusso nell'ultimo paragrafo del capitolo precedente.

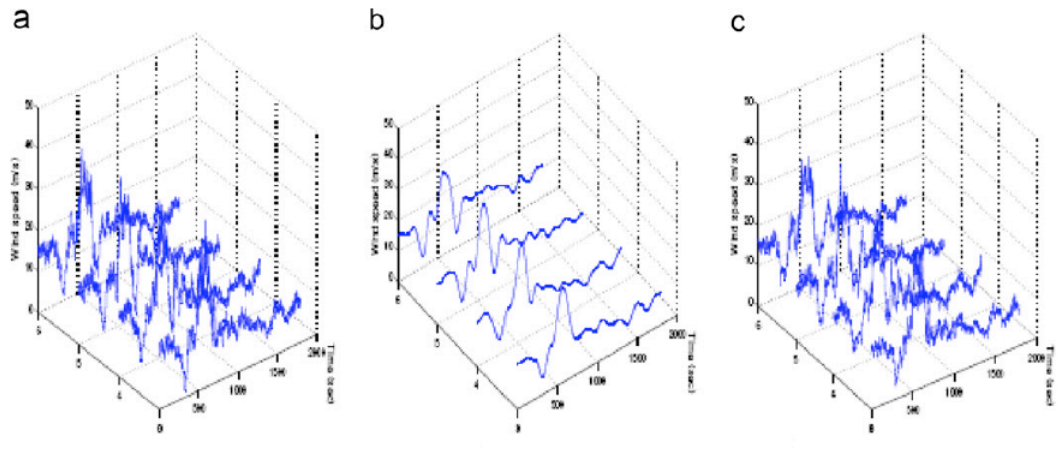


Figura 3.1: Processi stocastici e processi lineari

Senza perdita di generalità possiamo fissare $\psi_0 = 1$ cosicché

$$\begin{aligned} W_t &= \left(\psi_0 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \cdots\right) a_t \\ &= \Psi\left(B\right) a_t \end{aligned} \tag{3.2}$$

I pesi $\psi_t, t = 1, 2, \ldots,$ sono costanti (sebbene incogniti) e compaiono esplicitamente e con potenza uno; quindi il processo stocastico diventa lineare. Nella (3.1) si è posto, per semplicità, $\mu = 0$ anche se, in molte occasioni, tale parametro dovrà essere stimato in quanto potrebbe avere un significato specifico nel contesto applicativo della serie storica.

Il polinomio $\Psi(B)$ si chiama funzione di trasferimento (oppure filtro) perché trasmette (o filtra) l'effetto degli errori (detti anche correzioni o innovazioni) a_t nei valori della serie storica W_t. Il polinomio $\Psi(B)$ è anche noto come equazione caratteristica o polinomio caratteristico del processo stocastico lineare.

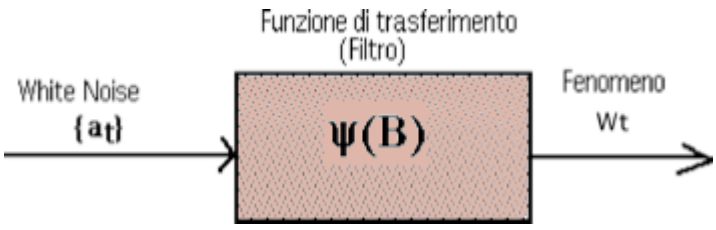


Figura 3.2: Effetto del polinomio caratteristico

La funzione di trasferimento è quella che definiamo la *black barrack* in cui un processo stocastico di eventi indipendenti, imprevedibili, di per sé forse trascurabili, passando attraverso di esso, abbia generato la serie storica finita a cui ci troviamo di fronte. Quindi la *barrack* contiene quanto di importante possiamo conoscere sul nostro processo.

E' essenziale evidenziare il presupposto che auto-rizza a ritenere che una serie erratica possa servire a modellare una serie empirica che invece riteniamo tutt'altro che casuale. Una buona parte della teoria poggia sul teorema della scomposizione di Wold, formulato nel 1954, per il quale si afferma che ogni processo stocastico stazionario, almeno in senso debole, possa essere descritto come la somma di due componenti separate:

$$W_t = X_t + Y_t \tag{3.3}$$

dove la parte X_t è una funzione deterministica ciclo-trend (cioè conoscibile e prevedibile senza errore) che varia nel tempo:

$$X_t = \mu + \sum_{j=1}^{\infty} \left[\alpha_j \sin\left(\delta_j t\right) + \beta_j \cos\left(\delta_j t\right)\right], \qquad 0 \leq \delta_j \leq \pi \tag{3.4}$$

La parte Y_t è espressa da una media mobile composta da infiniti termini di tipo *white noise* gaussiani:

$$Y_t = \sum_{t=0}^{\infty} \psi_t a_t \qquad a_t \sim N\left(0, \sigma^2\right) \tag{3.5}$$

Qui si possono solo fare affermazioni probabilistiche sui sui valori, passati, presenti e futuri. Sappiamo però che la somma dei pesi, in valore assoluto, è finita.

$$\sum_{t=0}^{\infty} |\psi_t| < \infty \tag{3.6}$$

Le due componenti, deterministica e stocastica, sono incorrelate $Cov\left(X_t, Y_t\right) = 0$. In generale, non è possibile stimare o verificare ipotesi sugli infiniti parametri ψ_t di W_t ed è quindi necessario approssimare W_t attraverso un processo con un numero contenuto di parametri (principio di parsimonia). Dal punto di vista delle previsioni, la parte deterministica del modello si può considerare come una componente stocastica *sui generis*, ma stabile per il periodo di tempo in cui il fenomeno rimane sotto la nostra attenzione.

Il teorema di Wold ha scarsa utilità pratica, ma è essenziale per trattare i processi stocastici in modo rigoroso.

Il processo stocastico lineare W_t può essere formulato in modo alternativo. Consideriamolo sia al tempo t che al tempo $t-1$

$$\begin{aligned} W_t &= a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \psi_3 a_{t-3} + \psi_4 a_{t-4} \cdots \\ W_{t-1} &= a_{t-1} + \psi_1 a_{t-2} + \psi_2 a_{t-3} + \psi_3 a_{t-4} + \psi_4 a_{t-5} \cdots \end{aligned} \tag{3.7}$$

Moltiplichiamo la seconda equazione per ψ_1

$$\begin{aligned} W_t &= a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \psi_3 a_{t-3} + \psi_4 a_{t-4} \cdots \\ \psi_1 W_{t-1} &= \psi_1 a_{t-1} + \psi_1^2 a_{t-2} + \psi_1\psi_2 a_{t-3} + \psi_1\psi_3 a_{t-4} + \psi_1\psi_4 a_{t-5} \cdots \end{aligned} \tag{3.8}$$

e sottraiamola dalla prima equazione

$$W_t - \psi_1 W_{t-1} = a_t + \left(\psi_2 - \psi_1^2\right) a_{t-2} + \left(\psi_3 - \psi_2\psi_1\right) a_{t-3} + \left(\psi_4 - \psi_3\psi_1\right) a_{t-4} + \cdots \tag{3.9}$$

Come si vede dal lato destro della relazione è sparito il termine a_{t-1}.

Analizziamo ora il processo W_t al tempo $t-2$,

$$\begin{aligned} W_t - \psi_1 W_{t-1} &= a_t + \left(\psi_2 - \psi_1^2\right) a_{t-2} + \left(\psi_3 - \psi_2\psi_1\right) a_{t-3} + \left(\psi_4 - \psi_3\psi_1\right) a_{t-4} + \cdots \\ W_{t-2} &= a_{t-2} + \psi_1 a_{t-3} + \psi_2 a_{t-4} + \psi_3 a_{t-5} + \psi_4 a_{t-6} \cdots \end{aligned}$$

moltiplichiamone entrambi i lati della seconda per $(\psi_2 - \psi_1^2)$ ipotizzando che $\psi_2 \neq \psi_1^2$

$$\begin{aligned} W_t - \psi_1 W_{t-1} &= a_t + \left(\psi_2 - \psi_1^2\right) a_{t-2} + \left(\psi_3 - \psi_2\psi_1\right) a_{t-3} + \left(\psi_4 - \psi_3\psi_1\right) a_{t-4} + \cdots \\ \left(\psi_2 - \psi_1^2\right) W_{t-2} &= \left(\psi_2 - \psi_1^2\right) a_{t-2} + \left(\psi_2 - \psi_1^2\right) \psi_1 a_{t-3} + \psi_2 a_{t-4} \\ &+ \left(\psi_2 - \psi_1^2\right) \psi_3 a_{t-5} + \left(\psi_2 - \psi_1^2\right) \psi_4 a_{t-6} \cdots \end{aligned}$$

e sottraiamola, membro a membro, da $W_t - \psi_1 W_{t-1}$ così da ottenere

$$\begin{aligned} W_t - \psi_1 W_{t-1} - \left(\psi_2 - \psi_1^2\right) W_{t-2} = a_t + \left(\psi_3 - 2\psi_1\psi_2 + \psi_1^3\right) a_{t-3} + \\ \left(\psi_4 - 2\psi_1\psi_3 + \psi_1^2 + \psi_2\psi_1^2\right) a_{t-4} + \cdots \end{aligned} \tag{3.10}$$

In questo modo si è eliminato anche il termine a_{t-2}. Ripetendo le stesse opearazioni per tutti i periodi si ottiene

$$\begin{aligned} W_t - \pi_1 W_{t-1} - \pi_2 W_{t-2} - \pi_3 W_{t-3} - \cdots) &= a_t \\ W_t - \pi_1 B W_t - \pi_2 B^2 W_t - \pi_3 B^3 W_t - \cdots &= a_t \\ \left(\pi_0 - \pi_1 B - \pi_2 B^2 - \pi_3 B^3 - \cdots\right) W_t &= a_t \end{aligned} \tag{3.11}$$

in cui si è posto $\pi_0 = 1, \pi_1 = \psi_1, \pi_2 = (\psi_2 - \psi_1^2)$, etc. In forma sintetica, l'ultima relazione si rappresenta con

$$\Pi(B) W_t = a_t. \tag{3.12}$$

Il meccanismo della funzione di trasferimento risulta pertanto invertito. Nel filtro ora entra W_t e ne esce a_t. Nella *black barrack* è rimasto, come in un setaccio, il contenuto informativo proprio del processo studiato.

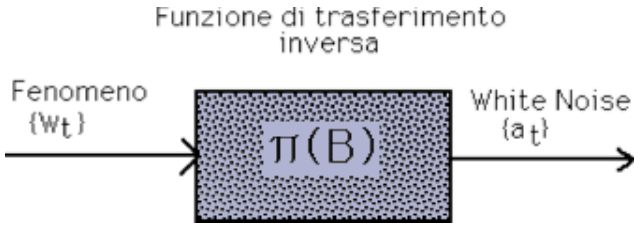


Figura 3.3: Filtro agente nella direzione opposta

Quello che della serie non è legato ai valori passati si ritiene sia costituito da *shocks* imprevedibili ed erratici.

Tra i due polinomi $\Psi(B)$ e $\Pi(B)$ che definisco le funzioni di trasferimento esiste una precisa relazione di dipendenza funzionale

$$W_t = \Psi(B)\, a_t; \quad \Pi(B)\, W_t = a_t \longrightarrow W_t = \Psi(B)\, \Pi(B)\, W_t \tag{3.13}$$

che può sussistere solo se $\Psi(B)\,\Pi(B) = 1$ ovvero $\Psi(B)$ e $\Pi(B)$ sono il reciproco l'uno dell'altro.

La struttura dei processi lineari è "semplice" sia in questa che nella formulazione precedente: fra le variabili casuali che formano il processo possono sussistere solo legami lineari. La semplicità della struttura permetterà, come vedremo nei prossimi paragrafi, di arrivare alla conoscenza del processo, o almeno alle conoscenze rilevanti avendo a disposizione una sola serie storica, cioè una sua sola realizzazione finita.

3.1.1 Funzione di auto-covarianza e di auto-correlazione per processi lineari

La funzione di auto-covarianza per i processi lineari è agevole da descrivere. Infatti

$$\begin{aligned} \gamma_k &= E(W_t W_{t-k}) \qquad (3.14)\\ &= E\left[(a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots)(a_{t-k} + \psi_1 a_{t-k-1} + \psi_2 a_{t-k-2} + \cdots)\right] \\ &= E\left[\sum_{r=0}^{\infty}\sum_{s=0}^{\infty} \psi_r \psi_s a_{t-r} a_{t-k-s}\right] = \sum_{r=0}^{\infty}\sum_{s=0}^{\infty} \psi_r \psi_s E(a_{t-r} a_{t-k-s}) \end{aligned}$$

Poiché

$$E(a_{t-r} a_{t-k-s}) = \begin{cases} 0 & \text{se } t-r \neq t-k-s \\ \sigma_a^2 & \text{se } t-r = t-k-s \end{cases} \tag{3.15}$$

consegue che

$$\gamma_k = \sigma_a^2 \left(\sum_{s=0}^{\infty} \psi_s \psi_{s+k} \right) \tag{3.16}$$

Se $\sigma_a^2 < \infty$ e la successione dei pesi ψ_s è finita oppure infinita, ma convergente, il processo W_t ha varianza finita. Questo si può verificare ponendo $k = 0$ nella (3.14)

$$\gamma_0 = \sigma_W^2 = \sigma_a^2 \left(\sum_{s=0}^{\infty} \psi_s^2 \right) \tag{3.17}$$

Quindi, perchè σ_W^2 sia finita è necessario che i pesi decrescano in valore assoluto piuttosto rapidamente, almeno a partire da un certo ritardo in poi.

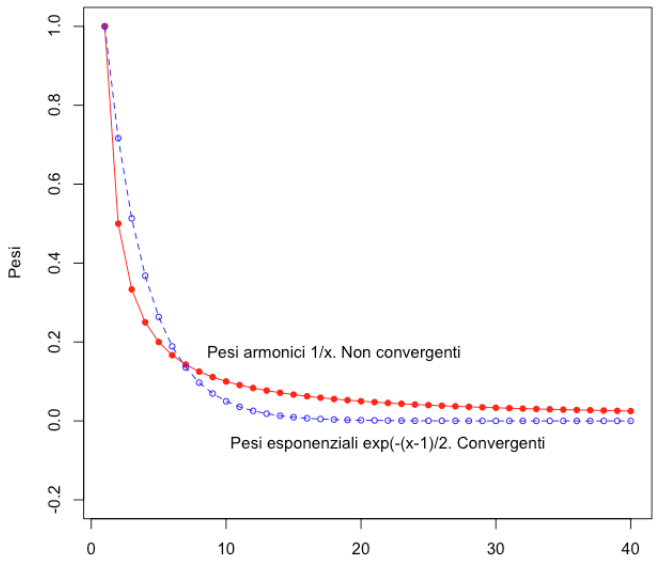


Figura 3.4: Andamento dei pesi

Le auto-correlazioni semplici dei processi lineari si ricavano dalla funzione di auto-covarianza

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\sum_{s=0}^{\infty} \psi_s \psi_{s+k}}{\sum_{s=0}^{\infty} \psi_s^2} \tag{3.18}$$

3.1.2 Stazionarietà ed invertibilità dei processi lineari

La stazionarietà è una condizione che individua quei processi stocastici in cui la distribuzione delle variabili casuali in essi incluse non si altera con il loro evolversi temporale o, almeno, non si alterano i loro momenti fino al secondo ordine (stazionarietà debole). In precedenza si è visto che le auto-covarianze di un processo stocastico stazionario obbediscono a dei vincoli intrinseci dovuti alla mera formulazione della matrice di varianze-covarianze (positività o semi-positività definita). Affinché anche

le γ_k di un processo stocastico lineare verifichino tali restrizioni bisogna e basta che la successione

$$\left(\psi_0 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \cdots\right) \tag{3.19}$$

sia convergente per ogni $|B| \leq 1$. In particolare, per $B = 1$, si richiede che sia

$$\sum_{s=0}^{\infty} \psi_s^2 < \infty \tag{3.20}$$

Tale convergenza si verifica, ad esempio, quando il numero di pesi non nulli è finito, oppure allorché i pesi decrescono "rapidamente" secondo il progredire del ritardo.

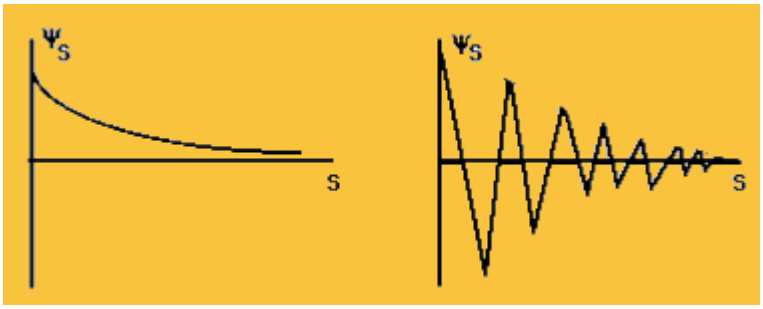


Figura 3.5: Pesi convergenti per un processo stocastico lineare

Se si realizza questo tipo di convergenza allora il processo stocastico lineare non solo ha varianza finita, ma è anche stazionario, dato che la convergenza per $|B|= 1$ implica la convergenza per $|B| \leq 1$.

L'invertibilità (od anche, reversibilità) è una caratteristica dei procesi stocastici lineari che fa riferimento alla possibilità di esprimere efficacemente il loro valore corrente in termini di un numero ridotto di loro osservazioni passate. Esaminiamo il processo stocastico seguente:

$$W_t = a_t - \theta a_{t-1} = (1 - \theta B)\, a_t \tag{3.21}$$

Il processo ricorda solo l'errore passato a_{t-1} ma, nel determinare il valore futuro, ammette un possibile errore coevo con W_t cioè a_t. La relazione (3.21) si ottiene troncando al primo termine il processo stocastico lineare (3.1). La funzione di trasferimento $\Psi(B) = (1 - \theta B)$ coinvolge ora il solo coefficiente θ e quindi processo è stazionario dato che la successione dei coefficienti in questo caso è finita e quindi convergente per costruzione. Inoltre, dalla identità basata sulla progressione

geometrica

$$(1-\theta B)\left[\sum_{s=0}^{k}(\theta B)^i\right]=\sum_{s=0}^{k}\left[(\theta B)^i-(\theta B)^{i+1}\right]=1-(\theta B)^{k+1}=1-\theta^{k+1}B^{k+1} \tag{3.22}$$

si ottiene

$$(1-\theta B)=\frac{1-\theta^{k+1}B^{k+1}}{\sum_{s=0}^{k}(\theta B)^i}=\frac{1-\theta^{k+1}B^{k+1}}{1+\theta B+\theta B^2+\cdots+\theta^k B^k} \tag{3.23}$$

per cui si avrebbe

$$\begin{aligned} W_t=(1-\theta B)\,a_t &= \frac{1-\theta^{k+1}B^{k+1}}{1+\theta B+\theta B^2+\cdots+\theta^k B^k}a_t \\ \left(1+\theta B+\theta B^2+\cdots+\theta^k B^k\right)W_t &= \left(1-\theta^{k+1}B^{k+1}\right)a_t \\ W_t+\theta W_{t-1}+\theta^2 W_{t-2}+\theta^3 W_{t-3}+\cdots+\theta^k W_{t-k} &= a_t-\theta^{k+1}a_{t-k-1}. \end{aligned}$$

Di conseguenza

$$W_t=a_t-\theta W_{t-1}-\theta^2 W_{t-2}-\theta^3 W_{t-3}+\cdots-\theta^k W_{t-k}-\theta^{k+1}a_{t-k-1} \tag{3.24}$$

Se $k\to\infty$, Il termine in $(k+1)$ si annulla prima degli altri per cui si arriva rapidamente alla linearità nel processo

$$W_t=a_t-\theta W_{t-1}-\theta^2 W_{t-2}-\theta^3 W_{t-3}+\cdots-\theta^k W_{t-k} \tag{3.25}$$

Qui, in particolare, vale la relazione $\psi_s=-\theta^s$. Il processo (3.25) è stazionario qualunque sia θ perché la somma non va indietro all'infinito, ma si interrompe al ritardo k-esimo. Tuttavia, se fosse $|\theta|\geq 1$ la dipendenza del valore corrente W_t da un valore precedente W_{t-k} aumenterebbe con l'ampliarsi del ritardo k e questo sarebbe contrario alla nostra ipotesi che i legami di dipendendenza siano forti solo per variabili ravvicinate nel tempo (processo ergodico). La coerenza con tale ipotesi si ottiene richiedendo che il processo sia invertibile ovvero richiedendo che la serie di potenze presente nel filtro $\Pi(B)$ che descrive il processo a_t in termini dei valori presenti e passati del processo (si guardi alla relazione (3.12)), ovvero si richiede che

$$\pi_0 B^0-\pi_1 B-\pi_2 B^2-\pi_3 B^3-\cdots- \tag{3.26}$$

sia convergente per ogni $|\theta|<1$. Come nel caso della stazionarietà, per la invertibilità si richiede che i pesi π decrescano, in valore assoluto, ed abbiano come limite lo zero, rapidamente raggiunto.

3.2 Processi auto-regressivi

L'espressione generale di questi processi è la seguente

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \phi_3 W_{t-3} + \cdots + \phi_p W_{t-p} + a_t \tag{3.27}$$

dove a_t è un processo *white noise*. Le serie storiche generate da (3.27) partono dal tempo $t = 1$ anche se si suppone che il processo sia già in atto da almeno p periodi. Il valore corrente W_t è determinato dalla somma ponderata dei p valori appena passati più un errore coevo a_t. La similarità con i modelli di regressione lineare multipla giustifica il termine "regressivi" nella loro denominazione; la parte "auto" è ovviamente dovuta al fatto che il modello utilizza solo valori interni al fenomeno rappresetato da W_t.

Un modo diverso di proporre il processo auto-regressivo di ordine p, indicato con AR(p), è il seguente

$$\begin{aligned} W_t - \phi_1 W_{t-1} - \phi_2 W_{t-2} - \phi_3 W_{t-3} - \cdots - \phi_p W_{t-p} &= a_t \\ W_t - \phi_1 B W_t - \phi_2 B^2 W_t - \phi_3 B^3 W_t - \cdots - \phi_p B^p W_t &= a_t \\ \left(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \cdots - \phi_p B^p\right) W_t &= a_t \end{aligned} \tag{3.28}$$

e quindi $\Phi_p(B) W_t = a_t$. Poichè il filtro $\Phi_p(B)$ comprende solo un numero finito di termini, il processo risulta sicuramente invertibile (cioè i pesi di $\Psi(B) = \Phi_p^{-1}(B)$) sono covergenti. Affinché sia anche stazionario, i coefficienti ϕ devono verificare alcune condizioni. Definiamo

$$W_t = \frac{a_t}{\Phi_p(B)} = \frac{a_t}{(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \cdots - \phi_p B^p)} = \frac{a_t}{(G_1 - B)(G_2 - B)\cdots(G_p - B)} \tag{3.29}$$

in cui le $G_i, i = 1, 2, \ldots, p$ sono le radici del polinomio $\Phi_p(B)$. Se per il fattore i-esimo al denominatore dell'ultima relazione in (3.29) risulta $|B/G_i| < 1$, allora tale fattore può essere sviluppato come una serie geometrica infinita.

$$\frac{1}{1 - \dfrac{B}{G_i}} = \left[1 + \frac{B}{G_i} + \left(\frac{B}{G_i}\right)^2 + \left(\frac{B}{G_i}\right)^3 + \cdots\right] \tag{3.30}$$

Lo stesso accade per tutti i fattori del denominatore di (3.29), per cui avremo

$$\begin{aligned} W_t &= \frac{a_t}{(G_1 - B)(G_2 - B)\cdots(G_p - B)} \\ &= \left[1 + \frac{B}{G_1} + \left(\frac{B}{G_1}\right)^2 + \left(\frac{B}{G_1}\right)^3 + \cdots\right]\cdots \\ &\quad \left[1 + \frac{B}{G_p} + \left(\frac{B}{G_p}\right)^2 + \left(\frac{B}{G_p}\right)^3 + \cdots\right] a_t \\ &= \left(1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \cdots\right) a_t \end{aligned}$$

e, in definitiva, $W_t = \Psi(B)\, a_t$. Il processo AR(p) può dunque essere considerato una aggregazione ponderata di un numero infinito di *white noise*.

La stazionarietà dei processi lineari richiede la convergenza della successione dei pesi ψ e questa, nel caso dei processi lineari auto-regressivi, si ottiene richiedendo che l'equazione caratteristica presente nel denominatore del filtro $\Psi_p(B)$

$$\left(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \cdots - \phi_p B^p\right) \tag{3.31}$$

abbia radici che siano in modulo maggiori dell'unità cioè radici ricadenti al di fuori del cerchio unitario (questa terminologia serve a ricordare che alcune radici potrebbero essere dei numeri complessi).

La funzione di auto-correlazionie di un processo AR(p) si ricava dalla relazione già stabilita neli paragrafi precedenti

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p} \tag{3.32}$$

che ha in genere un comportamento da studiare modello per modello. Meno incerta è la funzione di auto-correlazione parziale dell'AR(p)

$$\phi_k = \begin{cases} 0 & \text{se } k > p \\ \neq 0 & \text{se} k \leq p \end{cases} \tag{3.33}$$

che vede azzerare tutte le auto-correlazioni di ritardo superiore al p-esimo. I processi auto-regressivi hanno applicazioni privilegiate nei contesti in cui si dispone di dati molto copiosi quali l'ingegneria meccanica, elettronica, e nello studio dei titoli finanziari perché i molti dati consentono di arrivare a ritardi elevati; inoltre, i dati sono qui meno affetti da errore che tendono in genere a compromettere il rendimento dei modelli AR.

3.2.1 Processo AR(1)

La serie storica ha origine in un processo stocastico lineare auto-regressivo del primo ordine se per essa vale la relazione:

$$W_t = \phi_1 W_{t-1} + a_t \qquad (1 - \phi_1 B)\, W_t = a_t \tag{3.34}$$

L'equazione caratteristica del processo AR(1) ha una sola radice: $B = 1/\phi_1$ e quindi il processo è stazionario se e solo se $|\phi_1| < 1$. Da notare che AR(p) è sempre invertibile a causa del numero finito di valori passati che entrano nella formazione di W_t.

I processi auto-regressivi del primo ordine hanno sicuri riscontri nella realtà. Ad esempio, l'occupazione aggregata al mese corrente può essere vista come una frazione fissa di quella del mese passato (una frazione perché qualche disoccupato può avere ottenuto un lavoro) più un nuovo gruppo alla ricerca di un posto. Se la consistenza di questo secondo gruppo è assimilabile ad un processo *white noise* con media $\mu\,(1 - \phi_1)$, allora la sequenza cronologica della disoccupazione W_t può formare un processo AR(1). Infatti

$$W_t = \phi_1 W_{t-1} + \mu\,(1 - \phi_1) + a_t \tag{3.35}$$

La funzione di auto-correlazione parziale di un processo AR(1) presenta un valore significativo solo al primo ritardo. La funzione di auto-correlazione semplice decade esponenzialmente (se $\phi_1 > 0$) oppure procede per oscillazioni smorzate (se $\phi_1 < 0$). Nei processi AR(1) con valori elevati e negativi di ϕ_1 osservazioni superiori (inferiori) alla media tendono ad essere seguite da valori più piccoli (più grandi) della media. Se nel processo è presente un valore di livello μ che agisce da riferimento per l'andamento del fenomeno

$$W_t = \mu + \phi_1 W_{t-1} + a_t \quad \text{con } \phi_1 = 1 \text{ oppure } \phi_1 = -1 \tag{3.36}$$

allora μ è detto "deriva" ed il processo ha la sigla RWD cioè *random walk with drift* cioè passeggiata aleatoria con deriva. Possiamo riscrivere il processo RWD come

$$W_t - W_{t-1} = \mu + a_t \longrightarrow (1 - B)\, W_t = \mu + a_t \longrightarrow W_t = W_0 + \mu + a_t \tag{3.37}$$

dove W_0 è il valore del fenomeno al suo inizio. Dietro l'RWD c'è un trend lineare intorno al quale oscillano i valori del processo. E' per questo che il *random walk* non è stazionario, ma lo diventa immediatamente dopo la differenza prima. Da notare il comportamento non stazionario del processo *random walk* in cui la posizione raggiunta è quella in cui già si trova, a meno di un errore.

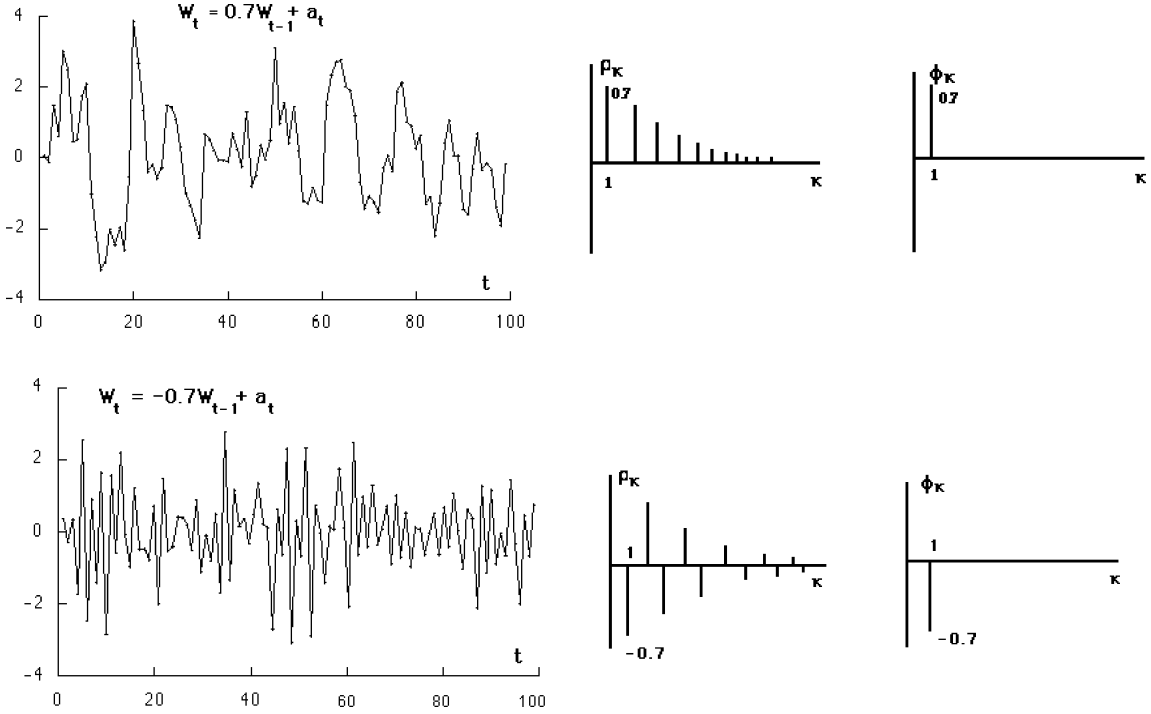


Figura 3.6: Processi auto-regressivi del primo ordine

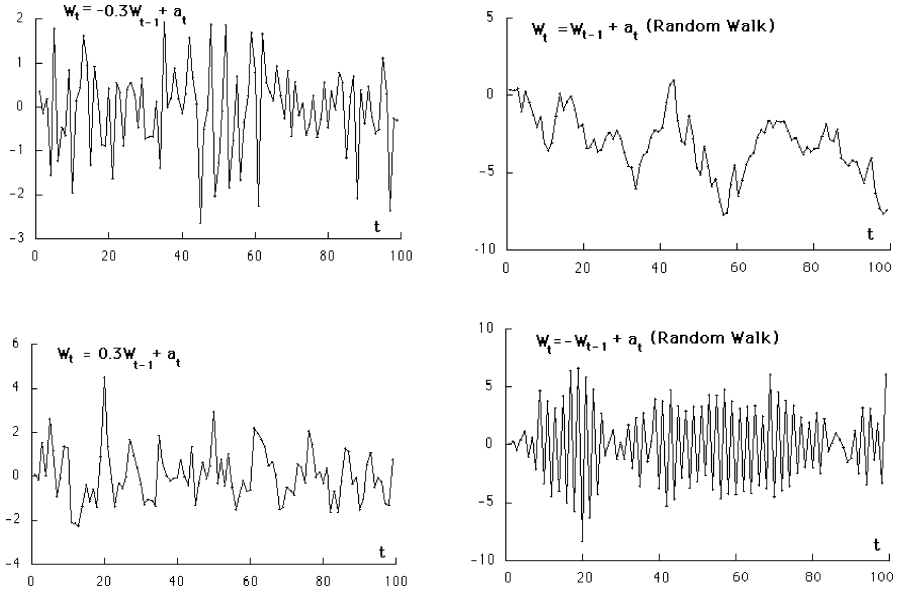


Figura 3.7: Processi quasi *white noise* e processo *random walk*

Nei processi AR(1) con valori moderati di ϕ_1 le oscillazioni aumentano se $\phi_1 > 0$ oppure diminuiscono se $\phi_1 < 0$. In entrambi i casi si realizza una tendenza al *white noise*; tendenza che si accentua al diminuire, in valore assoluto, del valore di ϕ_1. Nei processi AR(1) con valori molto elevati di ϕ_1 (negativi o positivi) il processo, pur rimanendo stazionario, mostra un comportamento simile a serie non stazionarie.

3.2.2 Processo AR(2)

Il processo AR(2) è definito dalla equazione

$$\begin{aligned} W_t - \phi_1 W_{t-1} - \phi_2 W_{t-2} &= a_t \\ \left(1 - \phi_1 B - \phi_2 B^2\right) W_t &= a_t \\ \phi_2 \left(B\right) W_t &= a_t \end{aligned} \tag{3.38}$$

Il processo AR(2) è stazionario se le radici della sua equazione caratteristica

$$\phi_2 \left(B\right) = 1 - \phi_1 B - \phi_2 B^2 = 0 \tag{3.39}$$

sono, in modulo, maggiori di uno. Ricordiamo che la locuzione "in modulo" scaturisce dal fatto che le radici della equazione caratteristica possono anche essere complesse e quindi, per la stazionarietà bisogna che stiano al di fuori del cerchio unitario (si veda la definizione di modulo di un numero complesso).

$$\left| \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2} \right| > 1 \tag{3.40}$$

Se il discriminante $\phi_1^2 + 4\phi_2 \geq 0$ le due radici sono reali e la funzione di auto-correlazione o decade esponenzialmente oppure procede per oscillazioni smorzate. Un comportamento simile ai processi AR(1) e da questi distinguibili in quanto le auto-correlazioni dell'AR(2) tendono a zero più lentamente. Se $\phi_1^2 + 4\phi_2 < 0$ le due radici sono complesse e la funzione di auto-correlazione mostra un caratteristico andamento sinusoidale. Se le radici sono, in modulo minori di uno, il modello AR(2) è stazionario. Tale condizione è soddisfatta se i parametri ϕ_1 e ϕ_2 ricadono nella regione triangolare di Stralkowski definito da

$$\gamma_k = \begin{cases} \phi_2 + \phi_1 < 1 \\ \phi_2 - \phi_1 < 1 \\ -1 < \phi_2 < 1 \end{cases} \tag{3.41}$$

Più lenta è la convergenza verso lo zero, più grande è numero di componenti auto-regressive che sarebbe utile considerare nella modellazione del processo stocastico. Processi AR di ordine molto elevato (ma non troppo) possono essere molto utili per modellare fenomeni a memoria lunghissima quali le serie idrogeologiche.

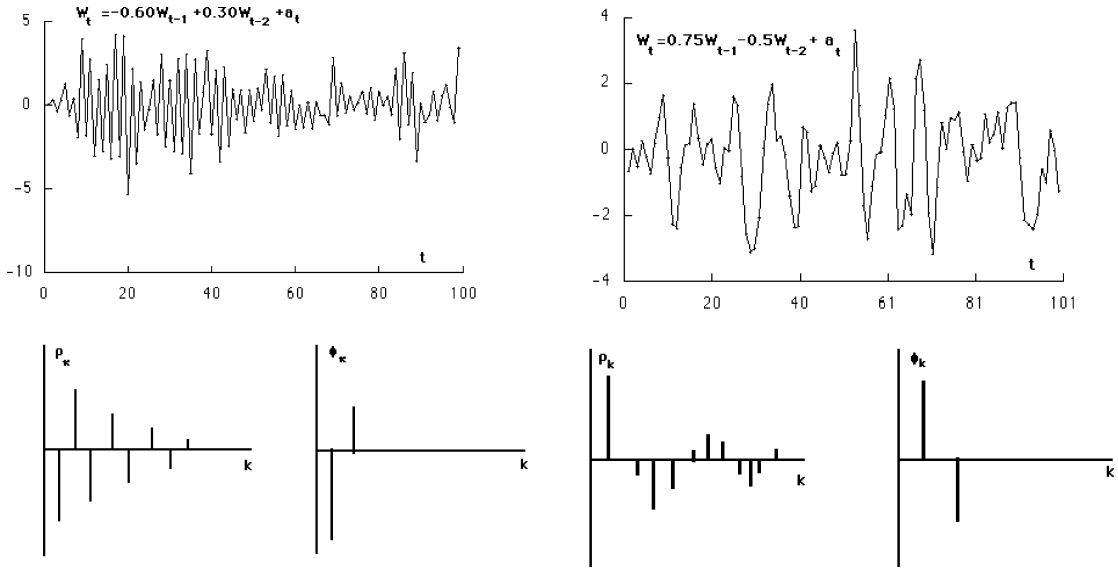


Figura 3.8: Esempi di AR(2)

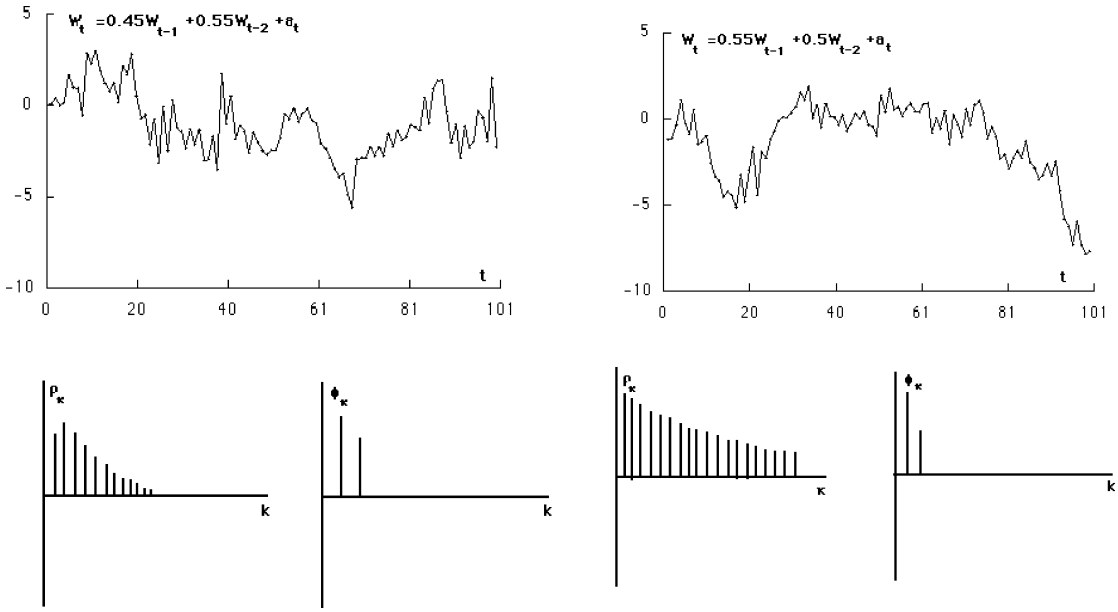


Figura 3.9: Altri esempi di AR(2)

3.3 Processi media mobile

L'espressione generale di quest classe di processi stocastici lineari è:

$$W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdot - \theta_q a_{t-q} \tag{3.42}$$

Il valore corrente W_t è dato dalla somma ponderata di errori casuali incorrelati avvenuti nei q periodi passati più un errore coevo. I processi media mobile, rappresentati con MA(q), possono insorgere in situazioni in cui un fenomeno forma il suo valore più recente reagendo ad una serie di *shocks* non del tutto assimilati in passato o che fanno sentire il proprio effetto per più periodi. Si pensi ad una stagione di lotta sindacale molto acuta per le serie economiche sul lavoro o ad un eccezionale periodo di tempo buono per le serie storiche della produzione agraria.

Un modo diverso di proporre il processo media mobile di ordine q è il seguente

$$W_t = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right) a_t \;\rightarrow\; W_t = \Theta_q(B) \tag{3.43}$$

Poiché il filtro $\Theta_q(B)$ comprende solo un numero finito di termini *white noise*, il processo risulta sicuramente stazionario. Affinché sia anche invertibile o reversibile, i coefficienti θ devono verificare alcune condizioni. Riscriviamo il processo in modo da isolare l'errore attuale

$$\begin{aligned} a_t = \frac{1}{\Theta_q(B)} W_t &= \frac{1}{(1 - \theta_1 B - \theta_2 B^2 - \cdot - \theta_q B^q)} W_t \\ &= \left(\pi_0 - \pi_1 B - \pi_2 B^2 - \pi_3 B^3 - \cdots -\right) W_t \\ &= \Pi(B) W_t \end{aligned} \tag{3.44}$$

in cui si sfruttano le proprietà della serie geometrica per risolvere il polinomio al denominatore. L' invertibilità dei processi lineari si ottiene, come si è già detto per i processi con infiniti termini, assicurando la convergenza della successione dei pesi $\pi_t, t = 0, 1, 2, \cdots,$. Tale convergenza, nel caso dei processi media mobile, si ottiene quando l'equazione caratteristica

$$\left(1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q\right) = 0 \tag{3.45}$$

ha radici, in modulo, maggiori dell'unità. C'è una sorta di dualità tra la condizione di invertibilità per i processi media mobile e la condizione di stazionarietà per i processi auto-regressivi. L'analogia tra i due tipi di processi è anche più ampia. Infatti, si è visto che i processi auto-regressi di ordine finito possono essere rappresentati da processi media mobile di ordine infinito (media ponderata di infiniti errori). Allo

stesso modo, processi media mobile di ordine finito possono essere espressi come processi auto-regressivi di ordine infinito.

Questo significa che i processi media mobile sono equivalenti a quelli auto-regressivi nel senso che se un modello del primo tipo è adatto a rappresentare il processo, lo sarà anche un modello del secondo tipo, sebbene con un numero di parametri molto più elevato o addirittura infinito. Dovendo scegliere il modello di processo più consono ad una serie storica osservata, saremo guidati dal principio della parsimonia: fra due modelli alternativi e di efficacia equivalente sceglieremo quello che prevede meno parametri incogniti.

La funzione di auto-covarianza di un processo MA(q) si ricava dalla relazione

$$E\left(a_t a_{t-k}\right) = E\left[\left(a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdot - \theta_q a_{t-q}\right)\right.$$
$$\left.\left(a_{t-k} - \theta_1 a_{t-k-1} - \theta_2 a_{t-k-2} - \cdot - \theta_q a_{t-k-q}\right)\right] \quad (3.46)$$

che implica una funzione di auto-covarianza del tipo

$$\gamma_k = \begin{cases} \left(-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_q\theta_{k+q}\right)\sigma_a^2 & \text{se } k \leq q \\ 0 & \text{se } k > q \end{cases} \quad (3.47)$$

dato che gli errori di epoche diverse sono incorrelate.

La funzione di auto-correlazione dell'MA(q) diventa

$$\rho_k = \begin{cases} \dfrac{\left(-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_q\theta_{k+q}\right)}{\left(1 + \theta_1^2 + \theta_2^2 + \theta_3^2 + \cdots + \theta_q^2\right)} & \text{se } k \leq q \\ 0 & \text{se } k > q \end{cases} \quad (3.48)$$

Ne consegue che le auto-correlazioni semplici dei processi MA(q) sono nulle per *lag* superiori a q e, quindi, il grafico delle auto-correlazioni semplici è quello più direttamente informativo per gli MA, mentre per gli AR tale ruolo lo aveva il grafico delle auto-correlazioni parziali. Se si sviluppa l'espressione di $\rho_k, k = 1, 2, \cdots, q$ in (3.48) si ottiene un sistema di equazioni che esprime i parametri incogniti $\boldsymbol{\theta}$ in funzione delle auto-correlazioni semplici. Tuttavia, a differenza delle equazioni Yule-Walker già usate per i parametri $\boldsymbol{\psi}$ dei processi auto-regressivi, il sistema presenta delle equazioni non lineari e quindi di non semplice soluzione. Inoltre, non è molto efficiente dal punto di vista statistico. Le soluzioni del sistema possono comunque essere usate per avere una idea della significatività dei coefficienti e per decidere quindi l'ordine del processo media mobile più compatibile con la serie storica osservata.

Lo schema delle auto-correlazioni parziali per il processo MA(q) è complessa e non se ne può dare una espressione generale in forma compatta e comprensibile. Ne vedremo invece le formule specifiche per i processi MA(1) ed MA(2). In generale

si può dire che i processi media mobile sono in contrapposizione con quelli auto-regressivi: laddove la funzione di auto-correlazione degli MA si azzera dopo il q-esimo *lag*, negli AR questo succede per le auto-correlazioni parziali. D'altra parte le auto-correlazioni parziali dei processi MA hanno un decadimento esponenziale (o procedono per oscillazioni smorzate); lo stesso comportamento, negli AR, si registra invece per le auto-correlazioni globali.

3.3.1 Processo MA(1)

Questo processo esprime il valore corrente W_t come proporzionale all'errore precedente (rispetto alla media zero o μ) a cui si somma l'errore attuale. Premesso che gli MA sono stazionari per costruzione, ecco alcune caratteristiche del processo MA(1)

$$\begin{aligned} \text{Formula :} \quad & W_t = a_t - \theta_1 a_{t-1} \\ \text{Equazione caratteristica :} \quad & (1 - \theta_1 B) \\ \text{Varianza :} \quad & \sigma_W^2 = \frac{\sigma_a^2}{1 + \theta_1^2} \\ \text{Invertibilità :} \quad & |\theta_1| < 1 \end{aligned}$$

Le funzioni di auto-correlazione semplice e parziale sono:

$$\rho_h = \begin{cases} \dfrac{-\theta_1}{1 + \theta_1^2} & \text{se } k = 1 \\ 0 & \text{altrimenti} \end{cases} \qquad \phi_h = \frac{(-\theta_1)\,(1 - \theta_1^2)^k}{1 - \theta_1^{2(k+1)}} \tag{3.49}$$

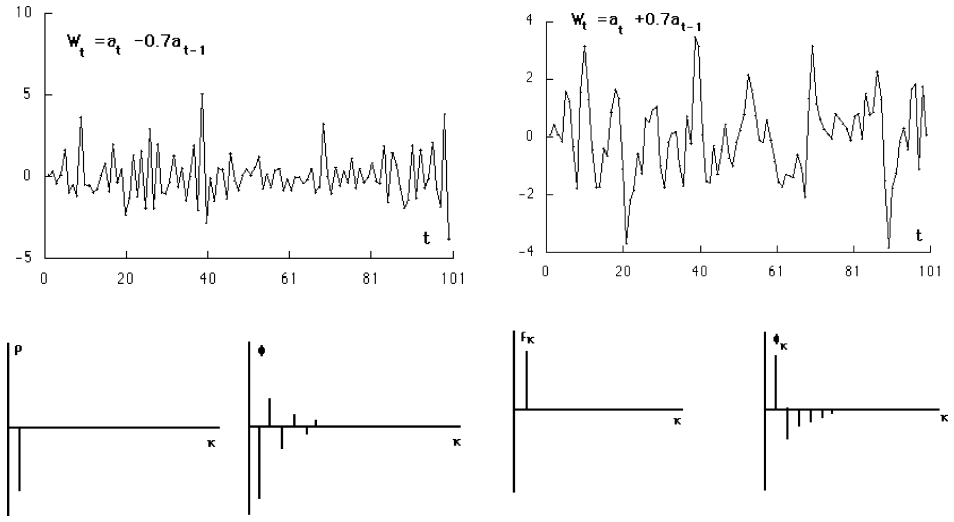


Figura 3.10: Esempi di processi MA del primo ordine

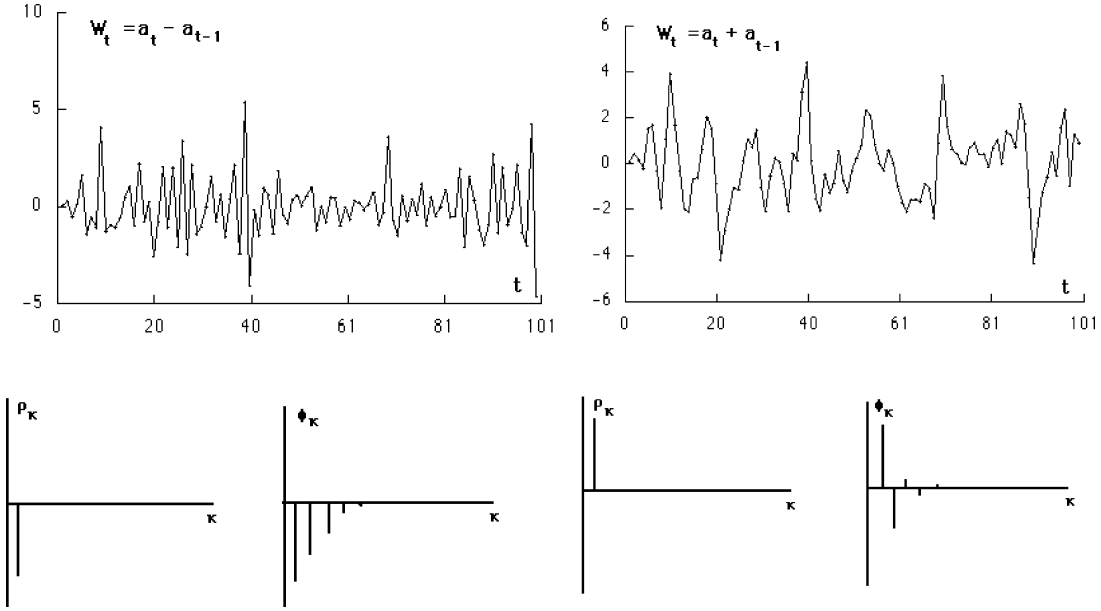


Figura 3.11: MA con radici unitarie e quindi non invertibili

Per meglio intendere il significato della invertibilità dei processi media mobile consideriamo i due processi

$$1^\circ) W_t = a_t - \theta a_{t-1}; \qquad 2^\circ) W_t = a_t - \frac{1}{\theta} a_{t-1} \tag{3.50}$$

I processi media mobile sono sempre esprimibili in termini dei valori passati della serie W_t. In particolare, per i processi MA(1) si ha

$$1^\circ)\ a_t = W_t + \theta W_{t-1} + \theta^2 W_{t-2} + \cdots + \quad 2^\circ)\ a_t = W_t + \frac{1}{\theta} W_{t-1} + \left(\frac{1}{\theta}\right)^2 W_{t-2} + \cdots + \tag{3.51}$$

Se $|\theta| < 1$, i valori passati del processo nel modello 1°) tenderebbero ad avere sempre meno influenza mentre il loro impatto diventerebbe esplosivo per il processo nel modello 2°). Tuttavia, dal punto di vista delle auto-correlazioni semplici i due processi sono equivalenti. Infatti

$$1^\circ)\ \rho_1 = \frac{-\theta}{1+\theta^2} \quad 2^\circ)\ \rho_1 = \frac{-\frac{1}{\theta}}{1+\left(\frac{1}{\theta}\right)^2} = \frac{-\theta}{1+\theta^2} \tag{3.52}$$

Lo studio del correlogramma non potrebbe, dunque, identificare univocamente il processo da cui è derivata la serie storica. Solo richiedendo che i processi siano invertibili possiamo, a partire dalle auto-correlazioni, risalire al processo stocastico.

3.3.2 Processo MA(2)

Qui si recepisce l'effetto delle due correzioni più recenti sul processo W_t. Anche in questo caso valori di W_t non intervengono direttamente ma attraverso una frazione dell'errore precedente e di quello subito indietro; a queste due correzioni, si somma l'errore attuale. Ecco alcune caratteristiche del processo MA(1)

$$\begin{aligned} \text{Formula :} \qquad & W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \\ \text{Equazione caratteristica :} \qquad & \left(1 - \theta_1 B - \theta_2 B^2\right) \\ \text{Varianza :} \qquad & \sigma_W^2 = \frac{\sigma_a^2}{1 + \theta_1^2 + \theta_2^2} \\ \text{Invertibilità :} \qquad & \theta_1 + \theta_2 < 1;\ \theta_2 - \theta_1 < 1;\ |\theta_2| < 1 \end{aligned}$$

Le funzioni di auto-correlazione semplice è data da

$$\rho_h = \begin{cases} \dfrac{-\theta_1 (1 - \theta_2)}{1 + \theta_1^2 + \theta_2^2} & \text{se } k = 1 \\ \dfrac{-\theta_2}{1 + \theta_1^2 + \theta_2^2} & \text{se } k = 2 \\ 0 & \text{altrimenti} \end{cases} \tag{3.53}$$

Se le due radici del polinomio al denominatore in (3.53) sono reali, la funzione di auto-correlazione parziale è dominata dalla somma di due andamenti esponenziali. Se invece le radici sono complesse, l'andamento è di tipo sinusoidale con onde di ampiezza decrescente cioè un comportamento opposto rispetto alle funzioni di auto-correlazione globale e parziale di un AR(2).

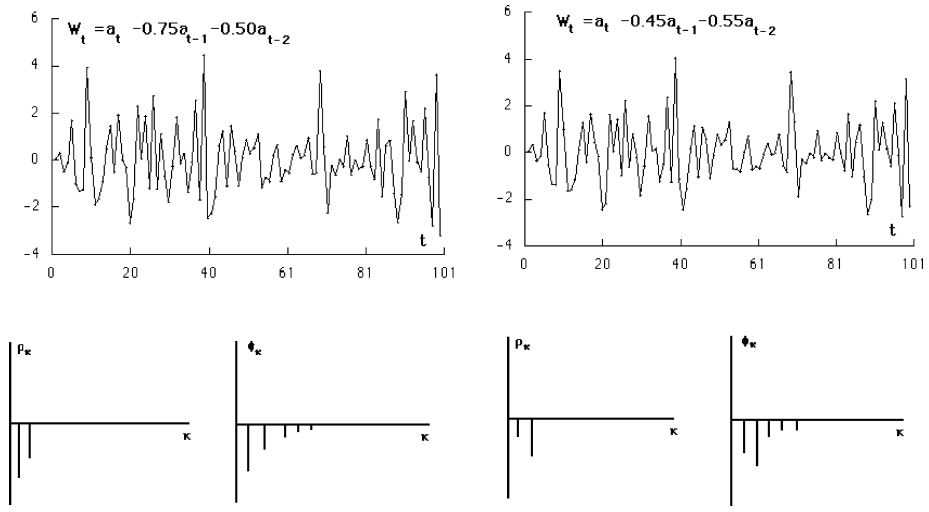


Figura 3.12: Esempio di MA(2)

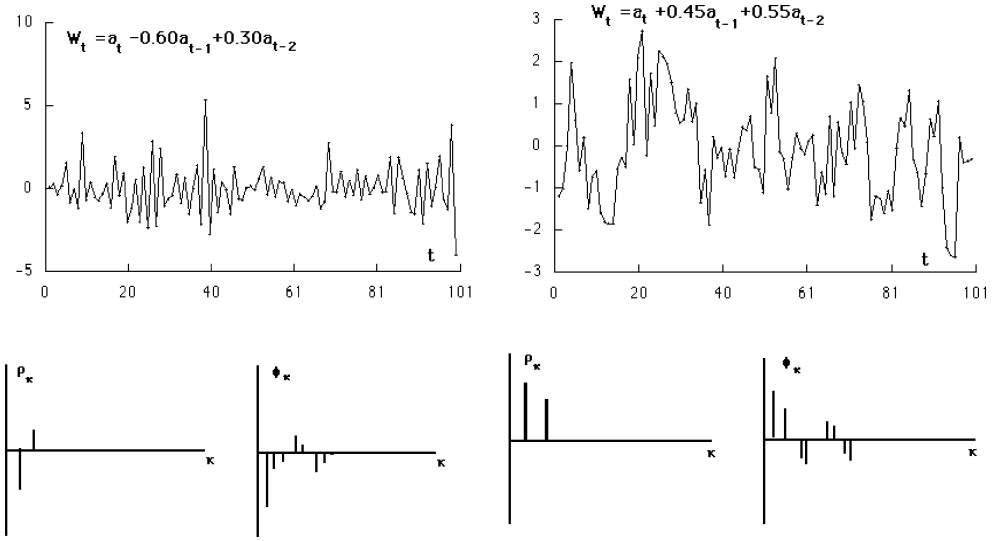


Figura 3.13: Esempio di MA(2)

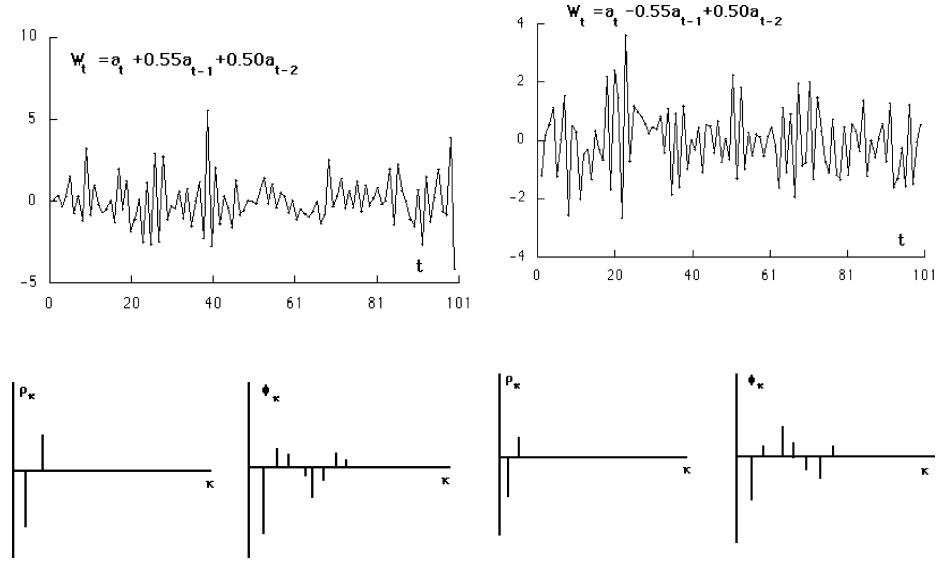


Figura 3.14: Esempio di MA(2)

3.4 Processi ARMA

Nei paragrafi precedenti sono stati esaminati dei processi che evidenziano i legami tra il valore corrente del fenomeno ed i suoi valori passati (processi AR) oppure con gli errori passati provenienti da un *white noise* (processi MA). Inoltre, è stato verificato che un processo con infiniti termini auto-regressivi AR(∞) può essere espresso come un media mobile di ordine finito e che un MA(∞) è equivalente ad un AR comprendente solo pochi termini. In estrema sintesi si è detto che sarà il principio di parsimonia ad orientare la scelta verso modelli con minori esigenze dal punto di vista della stima dei parametri incogniti.

Il principio di parsimonia nella rappresentazione dei processi può essere più pienamente realizzato da un modello che includa sia una componente media mobile che una componente auto-regressiva. Si propone in pratica la combinazione additiva di un processo AR(p) e di un processo MA(q) in un processo ARMA(p, q).

La formula generale di tali processi è

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \quad (3.54)$$

che vede un processo auto-regressivo (o di regressione con p variabili ritardate della dipendente) con errori descritti da una media mobile degli errori passati. La (3.54) può essere ri-espressa nei termini di due polinomi caratteristici

$$\begin{aligned} \left(1 - \phi_1 B - \phi_2 B^- \cdots - \phi_p B^p\right) W_t &= \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right) a_t \quad (3.55) \\ \Phi_p(B)\, W_t &= \Theta_q(B)\, a_t \end{aligned}$$

La stazionarietà del processo ARMA riguarda esclusivamente la sua componente auto-regressiva (la parte media mobile è stazionaria per definizione). Allo stesso modo l'invertibilità o reversibilità deve essere asseverata solo sulla componente media mobile, dato che la parte auto-regressiva è invertibile per definizione.

Quanto detto implica che il processo ARMA(p, q) è stazionario se e solo se le p radici del polinomio caratteristico della componente AR, $\Phi_p(B)$ sono, in modulo, maggiori di uno. In questo caso il processo ARMA(p, q) è equivalente ad un MA(∞)

$$\begin{aligned} W_t = \Phi^{-1}(B)\,\Theta_q(B)\, a_t &= \left[\frac{1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q}{1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p}\right] a_t \quad (3.56) \\ &= \left[\frac{(G_1 - B)(G_2 - B)\cdots(G_p - B)}{1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p}\right] a_t \end{aligned}$$

dove le $G_i, i = 1, 2, \ldots, p$ sono le radici del polinomio caratteristico della componente auto-regressiva. Se queste radici sono tali che $|G_i| > 1, i = 1, 2, \ldots, p$, ciascun fattore

$(G_i - B)$ può essere sviluppato in una serie geometrica convergente di modo che alla fine risulti

$$W_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \psi_3 a_{t-3} + \cdots \tag{3.57}$$

con i pesi $\psi_i \to 0$ per $i \to \infty$. Questa equivalenza può legittimare il ricorso ai modelli ARMA come più pratici rispetto ai modelli con sola componente MA.

Analizziamo, ad esempio, il processo misto ARMA$(1, 2)$

$$\begin{aligned} W_t &= \phi_1 W_{t-1} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \\ (1 - \phi_1 B) W_t &= \left(1 - \theta_1 B - \theta_2 B^2\right) a_t \\ W_t &= \left[\frac{-\theta_1 B - \theta_2 B^2}{1 - \phi_1 B}\right] a_t \end{aligned} \tag{3.58}$$

Se il processo è stazionario, o meglio, se lo è la sua parte auto-regressiva, allora $|G_1| = \left|\phi_1^{-1}\right| > 1$. Questo implica che i fattori al denominatore dell'espressione polinomiale dell'ARMA in (3.58) può essere sviluppato come una serie geometrica

$$\frac{1}{1 - \phi_1 B} = 1 + \phi_1 B + \phi_1^2 B^2 + \phi_1^3 B^3 + \cdots \tag{3.59}$$

La conseguenza sul modello ARMA$(1, 2)$ è che

$$\begin{aligned} W_t &= \left(1 + \phi_1 B + \phi_1^2 B^2 + \phi_1^3 B^3 + \cdots\right)\left(1 - \theta_1 B - \theta_2 B^2\right) a_t \\ &= \left[\left(1 + \phi_1 B + \phi_1^2 B^2 + \phi_1^3 B^3 + \cdots\right) + \right. \\ &+ \left(-\theta_1 B - \theta_1 \phi_1 B^2 - \theta_1 \phi_1^2 B^3 - \theta_1 \phi_1^3 B^4 + \cdots\right) + \\ &+ \left.\left(-\theta_2 B^2 - \theta_2 \phi_1 B^3 - \theta_2 \phi_1^2 B^4 - \theta_2 \phi_1^3 B^5 + \cdots\right)\right] a_t \\ &= \left[1 + (\phi_1 - \theta_1) B + \left(\phi_1^2 - \theta_1 \phi_1 - \theta_2\right) B^2 + \left(\phi_1^3 - \theta_1 \phi_1^2 - \theta_2 \phi_1\right) B^3 + \cdots +\right] a_t \end{aligned} \tag{3.60}$$

Se ora ridefiniamo i coefficienti in termini dei parametri di un processo MA puro

$$\psi_0 = 1,\ \psi_1 = \phi_1 - \theta_1,\ \psi_2 = \phi_1^2 - \theta_1 \phi_1 - \theta_2,\ \psi_i = \phi_1^i - \phi_1^{i-1} \theta_1 - \theta_2 \phi_1^{i-2} \tag{3.61}$$

otteniamo l'equivalenza del processo ARMA$(1, 2)$ con un MA(∞).

$$W_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} \tag{3.62}$$

Da notare che se si ipotizza la stazionarietà, si ha $|\phi_1| < 1$ e dunque i pesi tendono a zero quando aumenta il ritardo.

Il processo ARMA(p, q) è invertibile se e solo se le radici del polinomio caratteristico affererente alla componente media mobile $\Theta_q(B)$ sono, in modulo, maggiori di

uno. Se per la invertibilità si ripetono le operazioni già svolte per la stazionarietà, sarà facile accorgersi che un processo ARMA(p,q) invertibile è equivalente ad un processo AR(∞). Quindi, processi auto-regressivi o media mobile comprendenti un numero elevato di termini possano entrambi essere rappresentati da processi ARMA. Tale rappresentazione tende ad avere meno parametri delle rappresentazioni equivalenti dei due processi puri.

I modelli misti ARMA sono considerati con favore anche per il fatto che la somma di due processi auto-regressivi incorrelati genera un modello misto. Siano Y_t e Z_t due processi auto-regressivi di ordine p_y e di ordine p_z rispettivamente

$$\phi_{p_y}(B)\, Y_t = a_t; \qquad \phi_{p_z}(B)\, Z_t = b_t; \tag{3.63}$$

dove a_t e b_t sono due *white noise* mutualmente incorrelati. Il processo somma sarà

$$\begin{aligned} W_t = Y_t + Z_t &= \frac{a_t}{\phi_{p_y}(B)} + \frac{b_t}{\phi_{p_z}(B)} && (3.64)\\ &= \frac{a_t\phi_{p_z}(B) + b_t\phi_{p_y}(B)}{\phi_{p_y}(B)\,\phi_{p_z}(B)} \\ &= \frac{\theta(B)\, c_t}{\phi_{p_y}(B)\,\phi_{p_z}(B)} && (3.65) \end{aligned}$$

dove $\theta(B)\, c_t = a_t\phi_{p_z}(B) + b_t\phi_{p_y}(B)$ è un processo media mobile di un ordine uguale al massimo tra p_y e p_z. Questa proprietà torna molto utile se la serie storica in studio nasce da un aggregato di altre serie con comportamento auto-regressivo.

3.4.1 Auto-correlazioni semplici e parziali

Valutiamo il processo ARMA(p,q) al tempo $t-k$

$$\begin{aligned} W_{t-k} = \phi_1 W_{t-k-1} + \phi_2 W_{t-k-2} + \cdots + \phi_p W_{t-k-p} + a_{t-k} \\ - \theta_1 a_{t-k-1} - \theta_2 a_{t-k-2} - \cdots - \theta_q a_{t-k-q} \end{aligned} \tag{3.66}$$

Moltiplichiamo entrambi i lati della relazione per W_t e consideriamone il valore atteso delle risultanti espressioni

$$\begin{aligned} W_{t-k}W_t = (\phi_1 W_{t-k-1} + \phi_2 W_{t-k-2} + \cdots + \phi_p W_{t-k-p} + a_{t-k} \\ - \theta_1 a_{t-k-1} - \theta_2 a_{t-k-2} - \cdots - \theta_q a_{t-k-q})\,(W_t\phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} + a_t \\ -\theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}) \end{aligned} \tag{3.67}$$

$$\begin{aligned} E(W_{t-k}W_t) = \gamma_k = \phi_1\gamma_1 + \phi_2\gamma_2 + \cdots + \phi_p\gamma_p + \\ E(W_t a_{t-k}) - \theta_1 E(W_t a_{t-k-1}) - \theta_2 E(W_t a_{t-k-2}) - \cdots - \theta_q E(W_t a_{t-k-q}) \end{aligned} \tag{3.68}$$

dove $E(W_t a_{t-k})$ è la covarianza incrociata tra i valori del fenomeno e gli errori *white noise*. Poichè in W_t compaiono termini di errori a solo fino al tempo t ne consegue che

$$E(W_t a_{t-k}) = \begin{cases} 0 & \text{se } k > 0 \\ \neq 0 & \text{se} k \leq 0 \end{cases} \tag{3.69}$$

e quindi la funzione di auto-covarianza del modello ARMA(p, q) diviene

$$\gamma_k = \phi_1\gamma_1 + \phi_2\gamma_2 + \cdots + \phi_p\gamma_p \quad \text{per } k > q \tag{3.70}$$

Pertanto, la componente MA influenza le auto-covarianze fino al ritardo k e per valori superiori diventa simile alla auto-covarianza di un AR(p). Se nella relazione (3.68) si pone $k = 0$ si ottiene la varianza del processo.

$$\begin{aligned}\sigma^2_W = {} & \phi_1\gamma_1 + \phi_2\gamma_2 + \cdots + \phi_p\gamma_p + \\ & E(W_t a_{t-k}) - \theta_1 E(W_t a_{t-k-1}) - \theta_2 E(W_t a_{t-k-2}) - \cdots - \theta_q E(W_t a_{t-k-q})\end{aligned} \tag{3.71}$$

che non ha un espressione immediata come quella degli altri processi.

Discutiamo adesso il problema dei fattori comuni. Può capitare che i polinomi caratteristici delle due componenti abbiano una o più radici in comune. In questo caso il modello ARMA(p, q) equivale ad un altro ARMA di ordine più piccolo sia in p che in q. Consideriamo il modello ARMA$(2, 1)$

$$W_t = 0.2W_{t-1} + 0.15W_{t-2} + a_t + 0.3a_{t-1} \tag{3.72}$$

I polinomi caratteristici sono $\Phi_2(B) = 1 - 0.2B - 0.15B^2$ e $\Theta_1(B) = 1 + 0.3B$. Le radici del polinomio della parte AR sono 0.5 e -0.3 per cui il processo in (3.72) può essere riscritto come

$$W_t = \left[\frac{(1 + 0.3B)}{1 - 0.2B - 0.15B^2}\right] a_t = \left[\frac{(1 + 0.3B)}{(1 - 0.5B)(1 + 0.3B)}\right] a_t \tag{3.73}$$

che è in realtà un AR(1)

$$W_t = \left[\frac{1}{1 - 0.5B}\right] a_t \tag{3.74}$$

E' importante sottolineare che in questo caso sia l'ARMA$(2, 1)$ che l'AR(1) avrebbero la stessa funzione di auto-correlazione e sarebbero perciò indistinguibili. E' evidente che il principio della parsimonia ci porterà a preferire quello con meno parametri, ma non sempre la cancellazione dei termini è così nitida cosicché un controllo delle radici dei polinomi caratteristici è sempre opportuno. In questo senso, Anderson (1977), suggerisce una certa prudenza nell'ampliare il modello ARMA incrementando simultaneamente la parte AR e la parte MA dato che si potrebbero

determinare delle ridondanze nei parametri. Se, ad esempio, si prova ad applicare un ARMA$(1,1)$ ad un *white noise* ci si trova di fronte alla instabilità delle stime. Infatti, le due equazioni seguenti sono equivalenti

$$\begin{aligned} &1) && W_t = a_t \\ &2) && (1-\phi_1)\, W_t = (1+\theta_1)\, a_t \end{aligned}$$

per tutte le scelte tali che $\theta_1 = -\phi_1$.

3.4.2 Processo ARMA$(1,1)$

La combinazione di un processo AR(1) con un MA(1) produce un modello potente e con capacità di rappresentare serie storiche reali molto superiore a quanto ci si potrebbe aspettare dalla semplicità dei suoi componenti.

$$\begin{aligned} \text{Formula} &: W_t = \phi_1 W_{t-1} + a_t - \theta_1 a_{t-1} \\ \text{Equazione} &: W_t = \left[\frac{1-\theta_1 B}{1-\phi_1 B}\right] a_t \\ \text{Invertibilità} &: |\theta_1| < 1 \\ \text{Stazionarietà} &: |\phi_1| < 1 \\ \text{Varianza} &: \sigma_W^2 = \sigma_a^2 \left[\frac{1+\theta_1^2 - 2\theta_1\phi_1}{1-\phi_1^2}\right] \end{aligned}$$

La funzione delle auto-correlazioni semplici esiste in forma esplicita

$$\rho_k = \begin{cases} \dfrac{(1-\phi_1\theta_1)\,(\phi_1-\theta_1)}{1+\theta_1^2-2\phi_1\theta_1} & \text{se } k=1 \\ \rho_1 \phi_1^{k-1} & \text{se } k>1 \end{cases} \tag{3.75}$$

I pesi ψ per la rappresentazione come processo unico formato da infiniti errori/correzioni passate sono dati da

$$\psi_i = \begin{cases} \psi_1 = \phi_1 - \theta_1 & \text{se } k=1 \\ \psi_2 = \phi_1\psi_1 = (\phi_1-\theta_1)\,\phi_1 & \text{se } k=2 \\ \psi_k = \phi_1\psi_{k-1} = (\phi_1-\theta_1)\,\phi_1^{k-1} & \text{se } k>2 \end{cases} \tag{3.76}$$

I due sistemi di pesi possono essere usati per costruire sia un MA(∞) che un AR(∞). E' chiaro che questo non può implicare che ogni processo infinito MA o AR, possa essere ricondotto ad un ARMA$(1,1)$ sebbene accada spesso che AR o MA puri, ma

di ordine elevato, possano essere eguagliati in rendimento da un ARMA$(1,1)$. Le auto-correlazioni parziali tendono a zero senza mai annullarsi.

Nell'ARMA$(1,1)$ la funzione di auto-correlazione semplice dal *lag* 2 in poi non risente più della componente media mobile (che invece influenza la prima auto-correlazione) e presenta un comportamento simile a quello dell'AR(1). L'unica diversità è che nelle auto-correlazioni semplici dell'AR(1) la base elevata al ritardo è il semplice coefficiente ρ_1 laddove ARMA$(1,1)$ la base è ϕ_1 ottenuta da una formula che coinvolge anche la parte media mobile. Da notare infine che se $\phi_1 = \theta_1$ allora $\rho_1 = 0$ e quindi tutte le auto-correlazioni sarebbero nulle, ci si troverebbe in realtà non di fronte ad un processo ARMA$(1,1)$), ma di fronte ad un processo *white noise* ovvero un ARMA$(0,0)$.

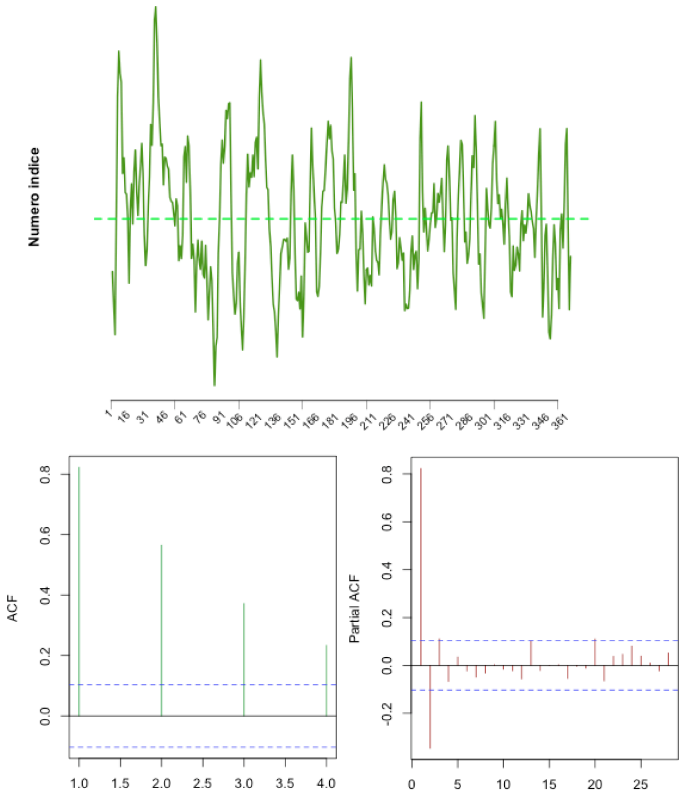


Figura 3.15: ARMA$(1,1)\,, \phi_1 = 0.5, \theta_1 = 0.5$

3.4.3 Processi ARIMA

Un processo che non sia stazionario in media può essere reso tale quando l'aspetto evolutivo è riconducibile ad un polinomio di un certo grado e cioè quando il ciclo-trend presente nella serie può essere approssimato con una espressione del tipo

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_d t^d \tag{3.77}$$

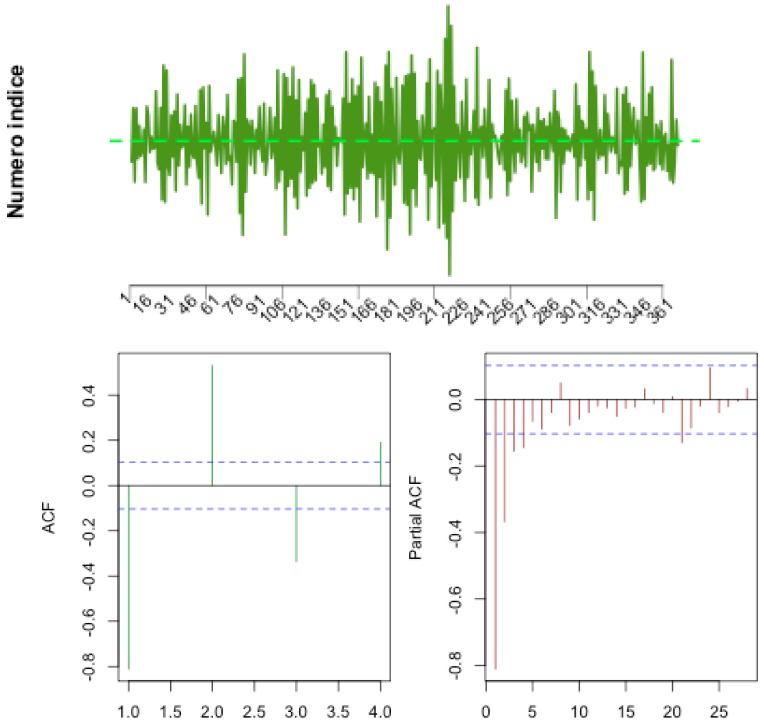


Figura 3.16: ARMA$(1,1)$, $\phi_1 = -0.7, \theta_1 = -0.5$

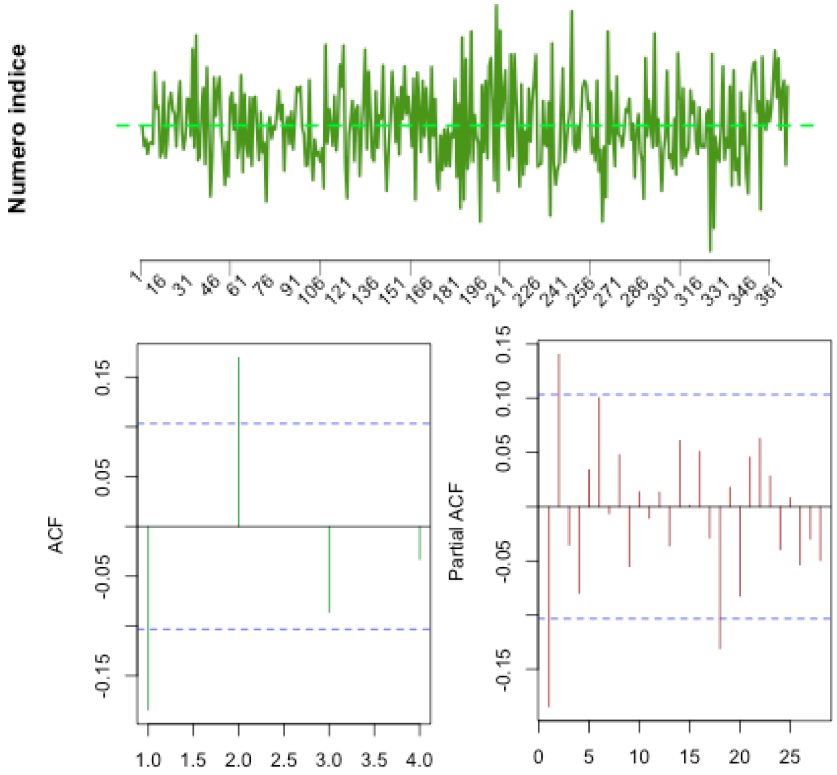


Figura 3.17: ARMA$(1,1)$, $\phi_1 = -0.7, \theta_1 = 0.5$

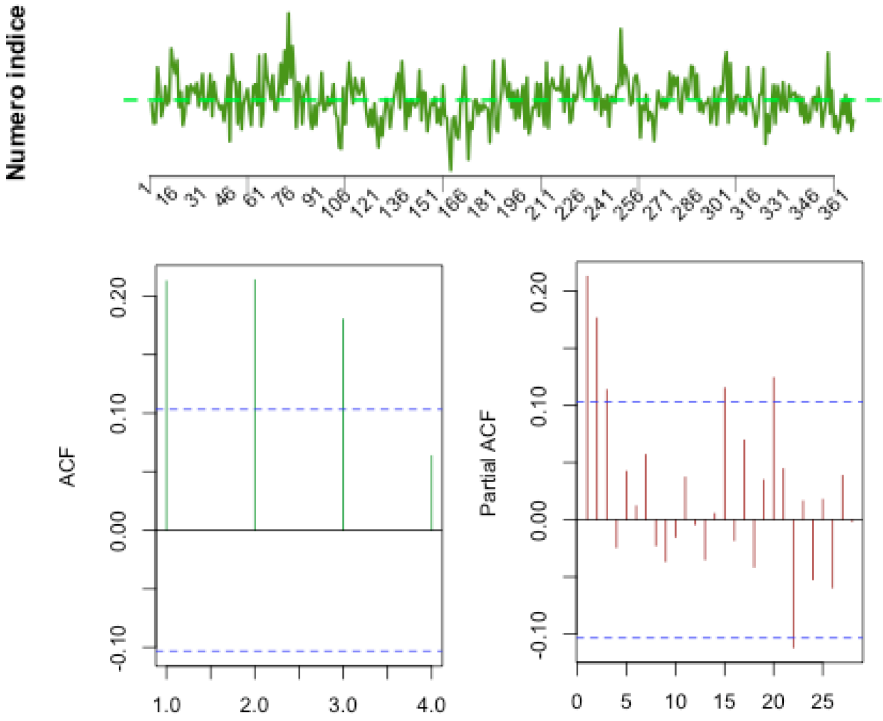


Figura 3.18: ARMA$(1,1)$, $\phi_1 = 0.7, \theta_1 = -0.5$

Questo argomento è stato già affrontato a proposito dell'applicazione dell'operatore differenza alla serie storica. Se W_t manca di stazionarietà, allora $\Delta = (1-B)^d W_t$ sarà un processo stazionario ARMA(p,q) qualora la serie includa un polinomio di grado d. Per dare conto della "I" nel nome di questi modelli occorre considerare che i processi passati attraverso il filtro $\Delta^d = (1-B)^d$ si dicono processi integrati ARMA(p,q) ovvero ARIMA(p,d,q). Infatti, se $\Delta^d X_t = W_t$ allora

$$X_t = \Delta^{-d} W_t = \left(1 + B + B^2 + B^3 + \cdots +\right) W_t \tag{3.78}$$

e quindi per riavere X_t occorre effettuare una somma (integrazione) del processo W_t.

Da un punto di vista teorico i processi ARIMA presentano una varianza infinita da cui discende una funzione di auto-correlazione non definita e covarianza legata al tempo oltre che al ritardo. Inoltre, la presenza del trend si riflette nella auto-correlazione calcolata su base campionaria con un andamento decrescente (esponenziale o oscillatorio) che si attenua con un andamento lento invece che rapido; in maniera simile, la funzione di auto-correlazione parziale tende a mantenersi vicino ad uno per diversi *lag*. La pratica di "detrendizzare" usando l'operatore differenza poggia sul presupposto che il ciclo-trend abbia natura determinista ovvero che nessuna sua componente erratica influenzi la parte ARMA di ciò che ne rimane dopo le differnze. Se si riscontra invece una certa volatilità, occorre modellare anche questa.

Quando la serie storica osservata W_t è un *white noise* una stima corretta della sua varianza è

$$\sigma_W^2 = \frac{(W_t - \mu_W)^2}{n-1}$$

La nostra premessa tuttavia è che W_t non sia del tutto casuale, ma abbia una componente deterministica più o meno accentuata. L'uso delle differenze, come si è visto nel paragrafo 2.7, è legittimato dall'ipotesi che nella parte deterministica agisca un fenomeno di ciclo-trend riconducibile ad un polinomio in t di un grado da determinarsi. Se questa ipotesi avesse fondamento, allora differenze successive dovrebbero ridurre la componente deterministica che dovrebbe attenuarsi a vantaggio della componente stocastica cosicché il rapporto tra le varianze

$$v_d = \frac{\sigma^2\left(\Delta^d W_t\right)}{\sigma^2\left(\Delta^{d+1} W_t\right)} \tag{3.79}$$

dovrebbe gradualmente avvicinarsi all'unità. Ne consegue che, il valore di d per cui v_d si approssima ad uno potrà essere considerato il giusto ordine di integrazione del processo ARIMA.

La differenziazione potrebbe introdurre nella serie storica a cui è applicata degli elementi estranei inducendo della correlazione aggiuntiva e artificiale. Nel software di supporto al presente testo abbiamo predisposto una indicazione euristica basata su un lavoro di [**?**]. Dopo aver calcolato la stima dello scarto quadratico medio della serie, diciamo σ_W, calcoliamo la regressione lineare avente come variabile dipendente la W_t e come regressori dei polinomi ortogonali di ordine d con d progressivamente crescente. Indichiamo con σ_d la stima dell'errore standard dei residui della regressione polinomiale anzidetta. Il primo valore per cui il rapporto σ_W/σ_d si avvicina ad uno indicherà in $(d-1)$ l'ordine delle differenze da valutare con più interesse. La tabella 3.1 mostra l'andamento della quantità v_d all'aumentare di d. L'ordine delle differenze suggerito è 2 o forse 4 dato che qui si realizza lo scarto più netto rispetto ai valori precedenti. Tuttavia, per $d = 4$ si realizzerebbe un intervento forte che

Tabella 3.1: Scelta dell'ordine delle differenze

d	v_d
1	0.657
2	0.512
3	0.237
4	0.239
5	0.043

deve essere motivato. C'è nel profilo della serie storica un polinomio di grado quarto e cioè due picchi e due valli? Se la risposta fosse negativa dovremmo optare per $d = 2$.

La figura che segue si riporta l'analisi grafica dell'effetto delle differenze sulle autocorrelazioni. Qui forse si può sciogliere il dubbio sul tipo di differenza da realizzare. Infatti a $d = 4$ la serie degli scarti comincia ad oscillare in modo improprio diversamente a quanto succede per $d = 2$ che risulta l'ordine di differenza più indicato.

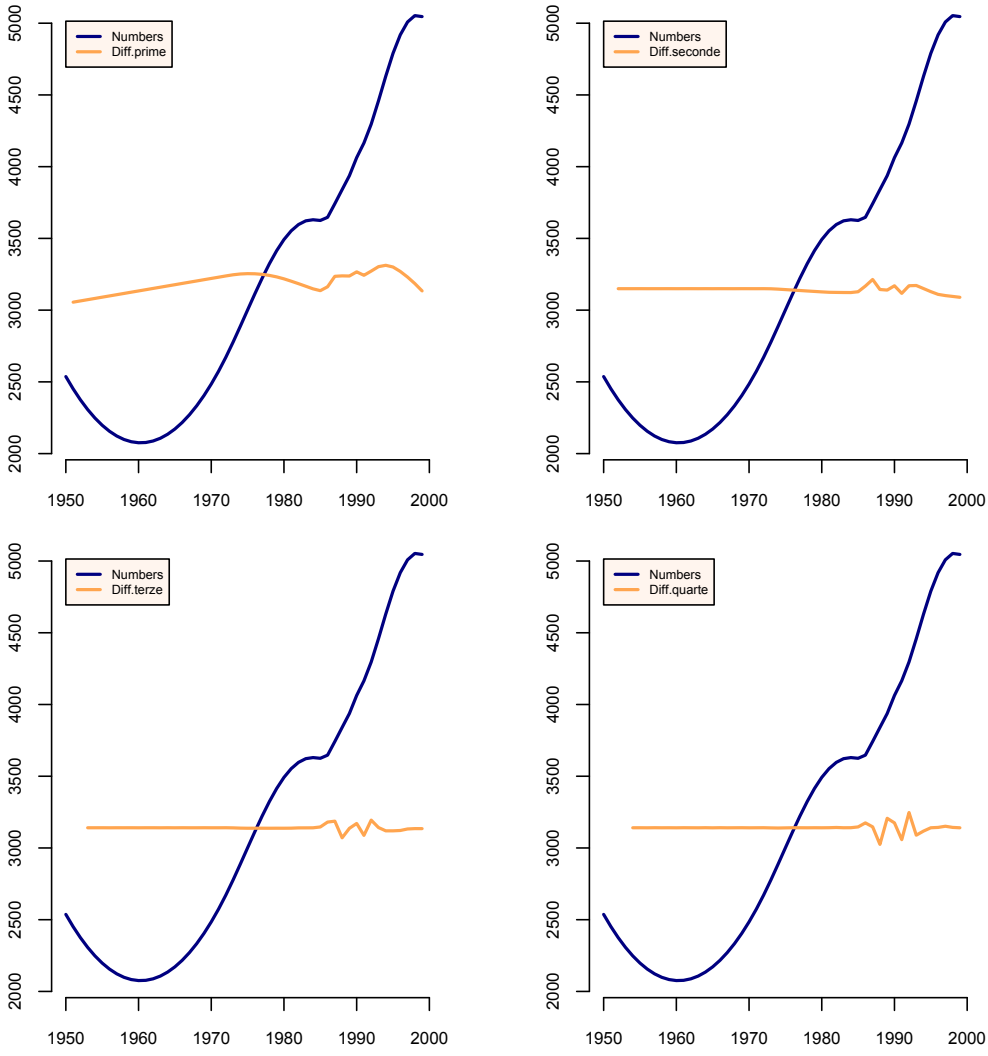


Figura 3.19: Number of Hospitals within the health care system

CAPITOLO 4

Modelli SARIMA

4.1 Introduzione

La stagionalità è costituita da movimenti periodici che si ripetono più o meno regolarmente nel corso dell'anno e da un anno all'altro. La definizione è vaga perché il problema della componente stagionale è complesso e di difficile trattazione: se non ci fosse stagionalità, analizzare le serie storiche sarebbe più semplice, ma è proprio nella stagionalità che risiede la parte più interessante dello studio.

Perché ci sia periodicità, lo stesso andamento si deve ripetere anno per anno, o meglio, la periodicità più rapida deve ripersi più o meno regolarmente in quella più lenta. Questo rende i movimenti stagionali riconoscibili... se presenti!. Infatti, la rilevazione periodica di un fenomeno non induce di per sé stagionalità: se per un arco di tempo prolungato si rileva il totale dei numeri estratti al Superenalotto in ciascuna estrazione e si scopre una periodicità è bene rivolgersi immediatamente alla magistratura (non penserete mica di utilizzarla?!).

La stagionalità può scaturire:

- Dal succedersi delle stagioni ed alle variazioni climatiche dovute alla rotazione terrestre intorno al sole: ad esempio le produzioni agricole o le costruzioni edilizie. Ai vecchi tempi, anche i matrimoni seguivano questi andamenti.
- Da convenzioni: la sospensione estiva della produzione o della attività scolastica. L'adozione dell'ora legale. La composizione dei mesi e delle settimane in termini di giornate lavorative (che variano in modo predeterminato). La scadenza di alcune tassazioni o del pagamento di dividendi.
- Da tradizioni: consumo di dolci nel periodo natalizio o pasquale.

- Da variazioni climatiche: consumo di combustibile. Vacanze scolastiche e ferie.

Per alcune serie storiche la periodicità è settimanale: la spesa in consumi alimentari è più bassa nell'ultima e più alta nella prima settimana del mese rispetto a quelle centrali. Le situazioni in cui "stagionalità" ed "annualità" diventano concetti da frammentare e articolare in subconcetti, sono molto diffuse:

(a) L'acquisto dei biglietti per il cinema ed il noleggio dei video, oltre a seguire i canoni delle stagioni climatiche, ha anche le periodicità infrasettimanali e quelle indotte da scadenze dei produttori che preferiscono lanciare i nuovi film o renderli disponibili per il mercato home in particolari periodi dell'anno.

(b) Le presenze negli alberghi e nelle località di vacanze che hanno svariati movimenti, sia stagionali che mensili, sia infrasettimanali che infra-giornalieri.

(c) Le serie barometriche, pluviometriche, termometriche idrografiche, che hanno varietà di comportamento con memoria a brevissimo e breve raggio, ma anche a medio e lungo termine. La modellazione di questi fenomeni è resa talvolta difficile dal comportamento asimmetrico rispetto agli aumenti rapidi seguiti da diminuzioni più lente.

(d) Le applicazioni mediche dell'analisi delle serie storiche stagionali hanno trovato terreno molto fertile ed ormai non si contano i casi di successo che hanno prodotto. Dalla rappresentazione dell'elettrocardiogramma al decadimento della presenza di un principio attivo nell'organismo,

(e) Serie stagionali si realizzano anche nel controllo della qualità dell'aria, dove l'unità di tempo può essere l'ora oppure i cinque minuti. L'irradiazione solare è analizzata per frazioni temporali di $3-5$ minuti. E' inevitabile che in queste serie storiche insorgano movimenti di varia frequenza.

(f) Nelle serie storiche sul traffico esiste un ciclo orario, giornaliero e settimanale.

(g) I due terzi dei costi di un *call center* deriva dal personale quindi prevedere il ritmo delle chiamate diventa uno strumento di controllo dei costi. In genere, la previsione settimanale o mensile può bastare.

Non sempre gli effetti di queste cause sono distinguibili ed anzi si sovrappongono. Nel nostro studio saremo costretti ad accantonare diverse componenti stagionali per ragioni di semplicità. Un caso emblematico è il consumo di energia elettrica che produce una serie storica composta di dati misurati ad intervalli di cinque minuti per 24 ore e sette giorni la settimana.

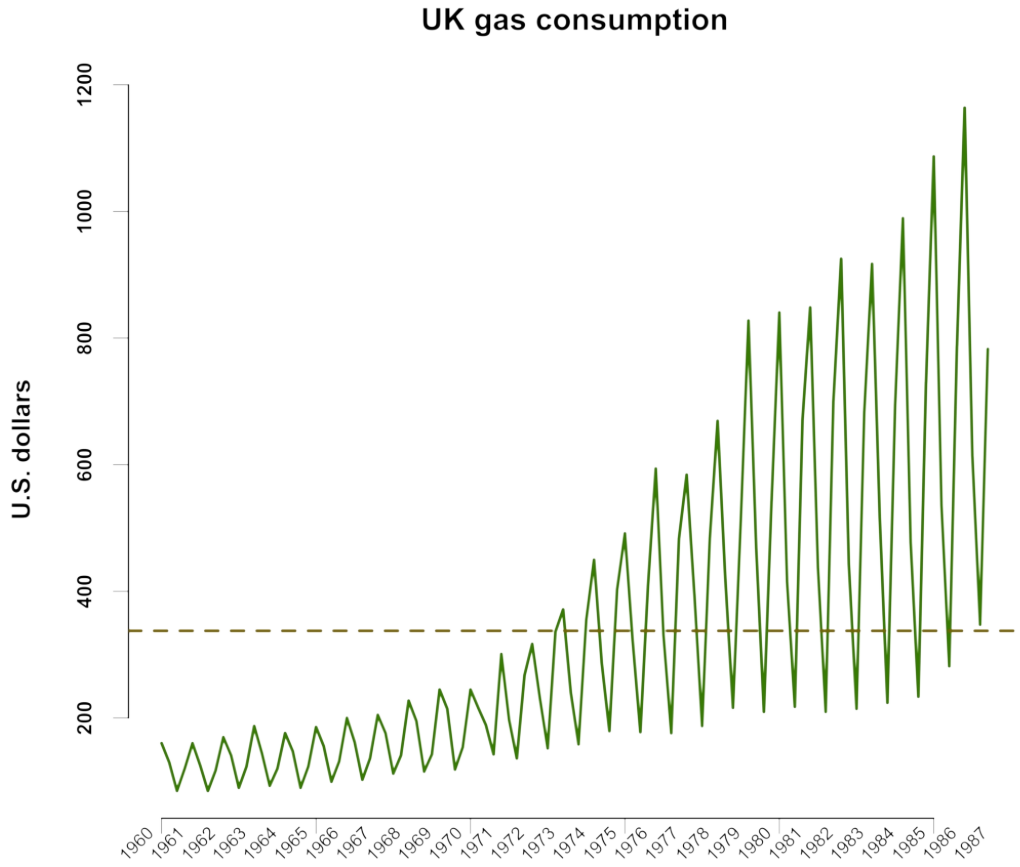


Figura 4.1: Organizzazione dei dati stagionali o periodici

Qui le periodicità si riscontrano innanzitutto all'interno della giornata nel senso che la domanda di energia è bassa nelle prime ore del mattino, cresce nel trascorrere della mattinata fino ad avere un piccolo calo all'inizio del pomeriggio; riprende a crescere per raggiungere il massimo picco all'inizio della serata. La richiesta rimane alta per poi diminuire progressivamente fino alla fase di valle nelle ore notturne. A questa periodicità si aggiunge quella settimanale che vede il consumo più elevato nei giorni centrali della settimana e ridotto nel week-end (il contrario succede però nelle località turistiche). C'è poi la variazione periodica tipica dell'evoluzione dei mesi con picchi in quelli in cui le temperature sono più estreme come in gennaio/agosto e valli in aprile/ novembre. E' ovvio che il fatto di ignorare alcune di queste cause non impedisca loro di agire comunque. L'approccio BJ è il nostro sistema di riferimento, ma può darsi che il problema sfugga a tale sistema anche se alcuni aspetti si accor-

dano bene con delle situazioni già note. In certi casi, per essere operativi, dovremo supporre che i legami importanti siano solo di tipo lineare e cioè rilevabili attraverso le auto-correlazioni semplici e parziali: non c'è alcuna garanzia che questo avvenga nella realtà dei nostri dati

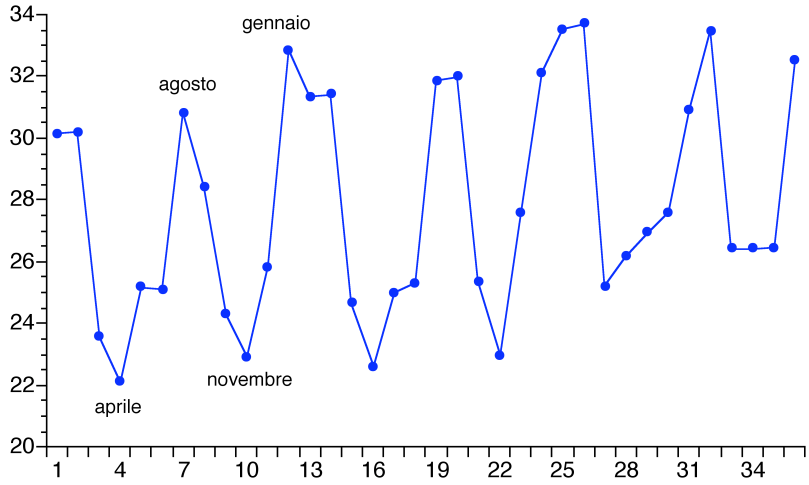


Figura 4.2: Esempio di serie storica stagionale sul consumo di energia

Nella figura 4.2 Le fluttuazioni più evidenti sono: un picco molto alto nei consumi del periodo estivo: luglio/agosto; un altro picco nei mesi invernali di dicembre e gennaio; riduzione dei consumi nelle mezze stagioni.

In numerose situazioni, l'atteggiamento verso le periodicità infra-serie è quello di rimuoverle ovvero destagionalizzare la serie storica nella convinzione che si tratti di un disturbo solo un po' meglio articolato dei tanti altri tipi di errore che affliggono i modelli statistici. Questo non è molto sensato se si ha come scopo una modellazione realistica dei dati: le componenti di una serie storica sono altamente intrecciate e non si può escluderne una senza eliminare - irrimediabilmente- una parte delle altre e non sempre quella più marginale. In questo senso, la metodologia Box-Jenkins si sforza di considerare la stagionalità esplicitamente anche pagando il costo connesso con l'aumento della complessità dei problemi che deve affrontare.

4.2 Arima Stagionali (SARIMA)

La estensione della metodologia BJ alle serie storiche stagionali può avvenire applicando un modello ARIMA $(P, D, Q)_s$ alle sottoserie periodiche, cioè un processo stocastico lineare specifico per ogni "stagione" che quindi, almeno potenzialmente,

è diverso per i diversi periodi. La notazione qui è stata leggermente modificata utilizzando le maiuscole per sottolineare il fatto che l'ambito di applicazione è ora quello stagionale, ma si è mantenuta la convenzione di indicare con la P la parte autoregressiva, la D per l'ordine di differenziazione e Q per la parte media mobile.

$$\Phi_P\left(B^s\right)\Delta_s^D W_t = \Theta_Q\left(B^s\right)a_t \tag{4.1}$$

dove

$$\Phi_P\left(B^s\right) = 1 - \Phi_1 B^s - -\Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps} \tag{4.2}$$

$$\Theta_Q\left(B^s\right) = 1 - \Theta_1 B^s - -\Theta_2 B^{2s} - \cdots - \Theta_Q B^{Qs} \tag{4.3}$$

$$\Delta_s^D = \left(1 - B^s\right)^D \tag{4.4}$$

L'idea è che l'effetto stagionale intervenga in tutti i periodi, ma con una portata che cambia da periodo a periodo e che tenda a seguire una struttura regolare. In altre parole, l'impatto stagionale si compensa sostanzialmente tra l'inizio e la fine del periodo di stagionalità: in una serie mensile, la somma degli effetti stagionali che cominciano a gennaio e terminano a dicembre, tenderanno a bilanciarsi, sia pure in modo incompleto nell'arco dell'anno.

Consideriamo la serie storica quadrimestrale descritta nella figura 4.3. Dovrebbe essere subito chiaro come non si possa affrontare uno studio di serie stagionali senza un numero adeguato di osservazioni. Nella serie riportata in sopra esistono 4 sotto-serie parziali di 27 elementi ciascuna ed una serie di $n = 108$ osservazioni globali.

	Qtr1	Qtr2	Qtr3	Qtr4					
1960	160.1	129.7	84.8	120.1	1974	449.9	286.6	179.3	403.4
1961	160.1	124.9	84.8	116.9	1975	491.5	321.8	177.7	409.8
1962	169.7	140.9	89.7	123.3	1976	593.9	329.8	176.1	483.5
1963	187.3	144.1	92.9	120.1	1977	584.3	395.4	187.3	485.1
1964	176.1	147.3	89.7	123.3	1978	669.2	421.0	216.1	509.1
1965	185.7	155.3	99.3	131.3	1979	827.7	467.5	209.7	542.7
1966	200.1	161.7	102.5	136.1	1980	840.5	414.6	217.7	670.8
1967	204.9	176.1	112.1	140.9	1981	848.5	437.0	209.7	701.2
1968	227.3	195.3	115.3	142.5	1982	925.3	443.4	214.5	683.6
1969	244.9	214.5	118.5	153.7	1983	917.3	515.5	224.1	694.8
1970	244.9	216.1	188.9	142.5	1984	989.4	477.1	233.7	730.0
1971	301.0	196.9	136.1	267.3	1985	1087.0	534.7	281.8	787.6
1972	317.0	230.5	152.1	336.2	1986	1163.9	613.1	347.4	782.8
1973	371.4	240.1	158.5	355.4					

Figura 4.3: Serie storica quadrimestrale sul consumo di metano nel Regno Unito

4.2.1 Differenziazione stagionale

Gli effetti stagionali possono essere fissi o variabili. Nel primo caso basta individuare il coefficiente che descrive l'effetto stagionale per ciascun elemento della stagionalità (ad esempio il coefficiente di ogni mese per le serie mensili o quello di ogni trimestre per le serie trimestrali). I coefficienti si ripeteranno fissi ed invariabili per tutto l'arco di osservazione della serie storica. Nel secondo caso occorre definire in che modo varia la stagionalità. Una possibilità è che gli effetti periodici più rapidi si propaghino secondo delle polinomiali che toccano le osservazioni in posizioni multiple della periodicità. Se questo è vero, possiamo rimuovere questa stagionalità con delle opportune differenze.

L'operatore differenza stagionale Δ_s^D si sviluppa per s, ma in modo diverso rispetto alle differenze regolari. Infatti, $\Delta = (1 - B^s) \neq (1 - B)^s$ perché

$$(1 - B)^s = \sum_{i=0}^{s} \binom{s}{i} (-1)^i B^i \tag{4.5}$$

con uno sviluppo di cui B^s è solo uno dei termini. L'effetto della differenza stagionale si può meglio comprendere considerando la seguente relazione che deriva dalla progressione geometrica

$$(1 - B^s) = (1 - B)\left(1 + B + B^2 + B^3 + \cdots + B^{s-1}\right) \tag{4.6}$$

in cui si affaccia l'operatore

$$A_s = \sum_{i=0}^{s-1} B^i = 1 + B + B^2 + B^3 + \cdots + B^{s-1} \tag{4.7}$$

che aggrega i valori successivi di una serie storica. Infatti:

$$A_s W_t = \sum_{i=0}^{s-1} B^i W_{t+i-1} = W_t + W_{t+1} + W_{t+2} + \cdots + W_{t+s-1} \tag{4.8}$$

Quindi,

$$\Delta_s W_t = (1 - B)\, A_s W_t = (A_s - B A_s)\, W_t = (A_s - A_{s+1})\, W_t \tag{4.9}$$

In breve, la differenza stagionale è una differenza tra due aggregati di stagioni successive. Applichiamo la differenza quadrimestrale alla serie riportata in figura 4.3 che, per $t = 10$, $s = 4$ e $D = 2$, comporta

$$\begin{aligned} \left(1 - B^4\right)^2 W_{10} &= \left(1 - 2B^4 + B^8\right) W_{10} \\ &= W_{10} - 2W_{10-4} + W_{10-8} \\ &= W_{10} - 2W_6 + W_2 \\ &= 140.9 - 2\,(124.9) + 129.7 \;=\; 20.8 \end{aligned} \tag{4.10}$$

N.B. Gli indici progrediscono dalla prima osservazione all'ultima in ordine di tempo con incremento unitario.

Le differenze stagionali devono essere effettuate con moderazione soprattutto se il numero di annualità per le quali si dispone dei dati è contenuto al punto da non consentire la identificazione di un *trend* anche quando è decritto da una formula molto semplice. Come nelle componenti non stagionali o regolari, le differenziazioni di ordine superiore allo stretto necessario non si limitano ad essere inutili, ma sono dannose dando l'ingannevole impressione di presenze strutturali *affascinanti, ma fasulle come la faccia di Megera*. Si veda la figura 4.4.



Figura 4.4: La duplice natura delle differenze non necessarie.

Un altro modo di vedere le differenze stagionali è il seguente. Supponiamo che la periodicità di interesse sia un numero pari, ad esempio $s = 4$. Ne consegue che

$$\Delta_4 = \left(1 - B^4\right) = \left(1 - B^2\right)\left(1 + B^2\right) = (1 - B)(1 + B)\left(1 + B^2\right) \tag{4.11}$$

L'operatore Δ_4 cattura la correlazione dei valori degli stessi trimestri quando è al livello massimo (cioè prossimo ad uno). Δ_4 ha quattro radici di cui due complesse: $+1, -1, \sqrt{-1}, -\sqrt{-1}$ che possono indurre un livello di differenziazione più ampio del dovuto.

Lo schema (4.10) è anche utile per mostrare una particolare caratteristica della modellazione di serie stagionali. Appare infatti in tutta evidenza come il *trend* quadratico stagionale si dipani nei soli valori del secondo trimestre della serie storica. Peraltro, non è difficile immaginare che se ci fosse un *trend* quadratico, anche diverso da trimestre a trimestre, le differenze seconde lo eliminerebbero senza lasciarne tracce apprezzabili. E fin qui il ragionamento regge. Vediamo quale potrebbe essere

lo svolgimento di un modello ARMA $(P,Q)_s$ con $t = 13$, $s = 4$, $P = 3$ e $Q = 2$ in cui la parte più marcata della periodicità sia stata rimossa con la differenziazione stagionale.

$$\begin{aligned} \left(1 - \Phi_1 B^4 - \Phi_2 B^8 - \Phi_3 B^{12}\right) W_{13} &= \left(1 - \Theta_1 B^4 - \Theta_2 B^8\right) a_{13} \quad (4.12) \\ W_{13} - \Phi_1 W_{13-4} - \Phi_2 W_{13-8} - \Phi_3 W_{13-12} &= a_{13} - \Theta_2 a_{13-4} - \Theta_1 a_{13-8} \quad (4.13) \\ W_{13} - \Phi_1 W_9 - \Phi_2 W_5 - \Phi_3 W_1 &= a_{13} - \Theta_2 a_9 - \Theta_1 a_5 \end{aligned}$$

$$W_{13} = \Phi_1 W_9 + \Phi_2 W_5 + \Phi_3 W_1 + a_{13} - \Theta_1 a_9 - \Theta_2 a_5$$

In questo caso i valori interessati sono i primi quattro della colonna $qdr1$ cioè le rilevazioni effettuate nel primo trimestre. Le nebbia si dirada: il modello ARMA di estrinseca nella sequenza definita dai multipli della periodicità: $\{1, 5, 9, 13, \ldots\}$. Questo è un punto accertato.

Ricordiamo lo scopo della metodologia BJ: determinare un modello di processo stocastico lineare che rappresenti l'evoluzione passata e futura della serie storica. Questo significa identificare l'ordine delle componenti, stimare i parametri del fattore autoregressivo e di quello media mobile in modo da far combaciare -per quanto possibile- la teoria sottostante il modello con la realtà dei dati osservati. Una strategia che è del tutto compatibile con l'idea che ci sia un distinto ARIMA $(P,D,Q)_s$ per ogni periodicità. In riferimento alla serie dei consumi di metano del Regno Unito, questo comporterebbe l'impegno a identificare quattro ARIMA $(P,D,Q)_s$, uno per trimestre, disponendo però di $n = 27$ dati per ciascuna sottoserie. La tabella esposta come figura 4.5 chiarisce l'aspetto computistico del processo.

	Trimestre (periodicità rapida)			
Anno (periodicità lenta)	Trim1	Trim2	Trim3	Trim4
1	W1	W2	W3	W4
2	W5	W6	W7	W8
…	…	…	…	…
t	Wt	Wt+1	Wt+2	wt+3
…	…	…	…	…

Figura 4.5: Organizzazione dei dati stagionali o periodici

La capacità degli elaboratori odierni rende l'aspetto dei calcoli poco rilevante anche se non è proprio trascurabile l'onere organizzativo: per una serie giornaliera occorrerebbe procedere alla identificazione, stima e validazione (si veda il capitolo 6) di 365 modelli diversi. E' vero, l'intero processo si può ridurre a qualche secondo

di calcolo con l'ausilio di un buon sistema esperto, ma dopo un notevole lasso di tempo per predisporre i programmi che lo guidano. Non è a questo altare che si può sacrificare la grande flessibilità di un ARIMA $(P, D, Q)_s$ multiplo. Infatti, la questione principale è proprio l'esiguo numero di dati che forma ciascuna delle sottoserie. Nella stragrande maggioranza dei casi, la numerosità ridotta non consentirebbe di avviare la benché minima analisi di una qualsiasi serie storica. Se invece si ipotizza che il modello ARIMA $(P, D, Q)_s$ sia lo stesso per tutti i periodi allora ogni sottoserie spinge i suoi dati a confluire in un unico processo che può essere quindi stimato grazie alla ergodicità. E' questa che consente di accomunare quattro sottoprocessi di $n = 27$ valori in un processo aggregato di $n = 108$ valori.

La componente stagionale della serie storica sarrà quindi descritta con il modello

$$\Phi_P(B^s)\,\Delta_s^D W_t = \Theta_Q(B^s)\,a_t \tag{4.14}$$

Analogamente a quanto si è detto per il modello ARIMA non stagionale, le condizioni di stazionarietà ed invertibilità del modello stagionale richiedono che le radici, rispettivamente di $\Phi_P(B^s)$ e $\Theta_Q(B^s)$ ricadano all'esterno del cerchio unitario.

Forse non è inutile aggiungere che esistono molte tecniche per affrontare la stagionalità dei dati e, nella classe ARIMA, se ne possono trovare di diverse in grado di proporre un adattamento più che accettabile dei dati osservati. Solo se però questi sono sufficientemente numerosi è possibile pensare seriamente ad individuare in modo convincente il modello più appropriato. In pratica, un centinaio di osservazioni potrebbero essere appena sufficienti per impostare una procedura realistica di identificazione e stima.

4.2.2 Esempi di modelli SARIMA

Uno dei modelli che si sono rivelati più utili per modellare serie storiche con componente stagionale è l'ARIMA $(0, 0, 1)_s$ che per $s = 12$ ha equazione:

$$W_t = a_t - \Theta_1 a_{t-12} \tag{4.15}$$

ed è descritto dalla seguente funzione di autocorrelazione:

$$\rho(\pm sk) = \begin{cases} \dfrac{-\Theta_1}{(1+\Theta_1^2)} & \text{per } k = 1 \\ 0 & \text{per } k \neq 1 \end{cases} \tag{4.16}$$

Nella figura 4.6 troviamo verificate le strutture di **acf** e di **pacf** tipiche del processo media mobile non integrato con un ritardo applicato ai valori che si intervallano di

$s = 12$ unità di tempo. In particolare, si può notare una singola autocorrelazione significativa ed il decadimento delle autocorrelazioni parziali. Plosser (1969) nello

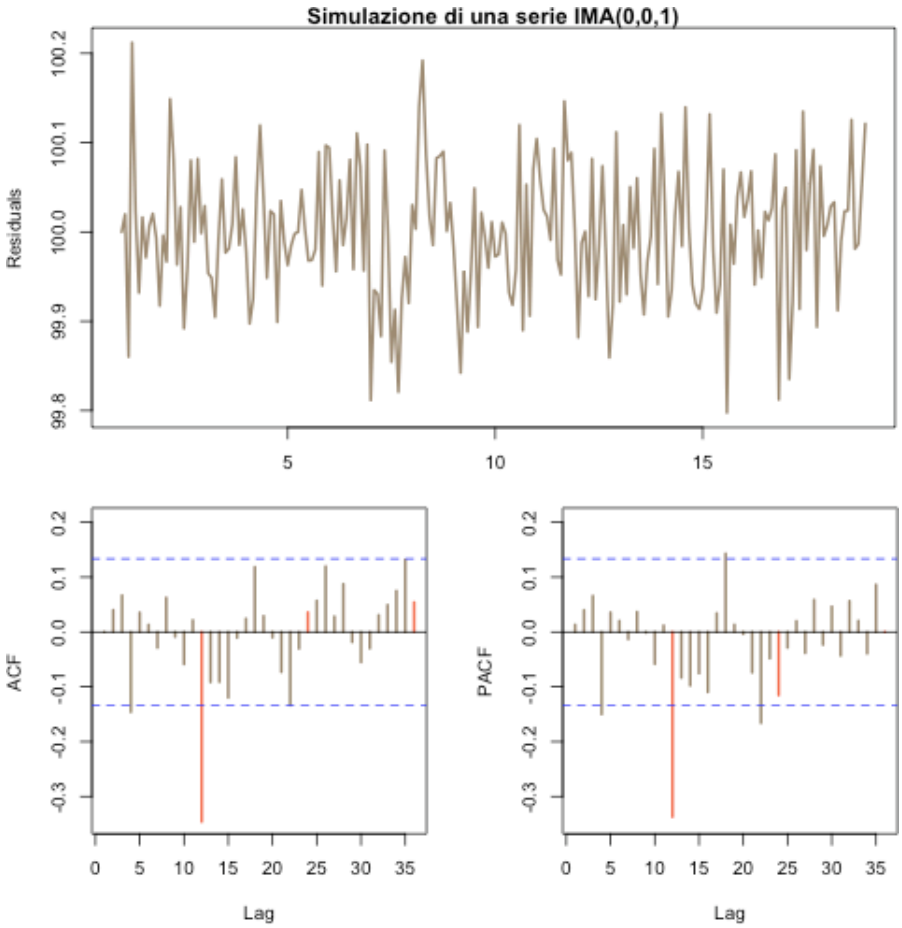


Figura 4.6: Esempio di processo con schema $\text{IMA}(0,0,1)_{12}$

studiare cinque serie storiche relative al mondo degli affari descrive molto bene la stagionalità delle serie solo con modelli di tipo $\text{IMA}(0,1,1)_{12}$ che includono anche un trend.

I modelli ARIMA stagionali sono anche definiti adattivi perché consentono alla stagionalità di modificare forma nel corso del tempo pur restando nella stessa classe. In altre parole, i SARIMA sono modelli a stagionalità variabile ben diversi dai coefficienti di stagionalità adoperati per filtrare gli effetti stagionali nell'approccio classico. Non è però ancora del tutto dimostrato che siano più efficienti delle serie di coefficienti a stagionalità variabile polinomialmente.

Un modello stagionale, pure ricorrente, è quello che si limita a coinvolgere un singolo termine autoregressivo $\text{ARI}(1,0,0)_s$ che per $s = 12$ ha equazione:

$$W_t = \Phi_1 W_{t-12} + a_t \tag{4.17}$$

Il modello è particolarmente semplice ed è descritto dalla funzione di autocorrelazione:

$$\rho(\pm 12k) = \begin{cases} \Phi_1^k & \text{per } k = 0, 1, 2, 3, \cdots \\ 0 & \text{altrimenti} \end{cases} \tag{4.18}$$

Nella figura 4.7 notiamo il decadimento delle autocorrelazioni semplici (per lag di 12 in 12) e le autocorrelazioni parziali che sono significative solo al lag 12.

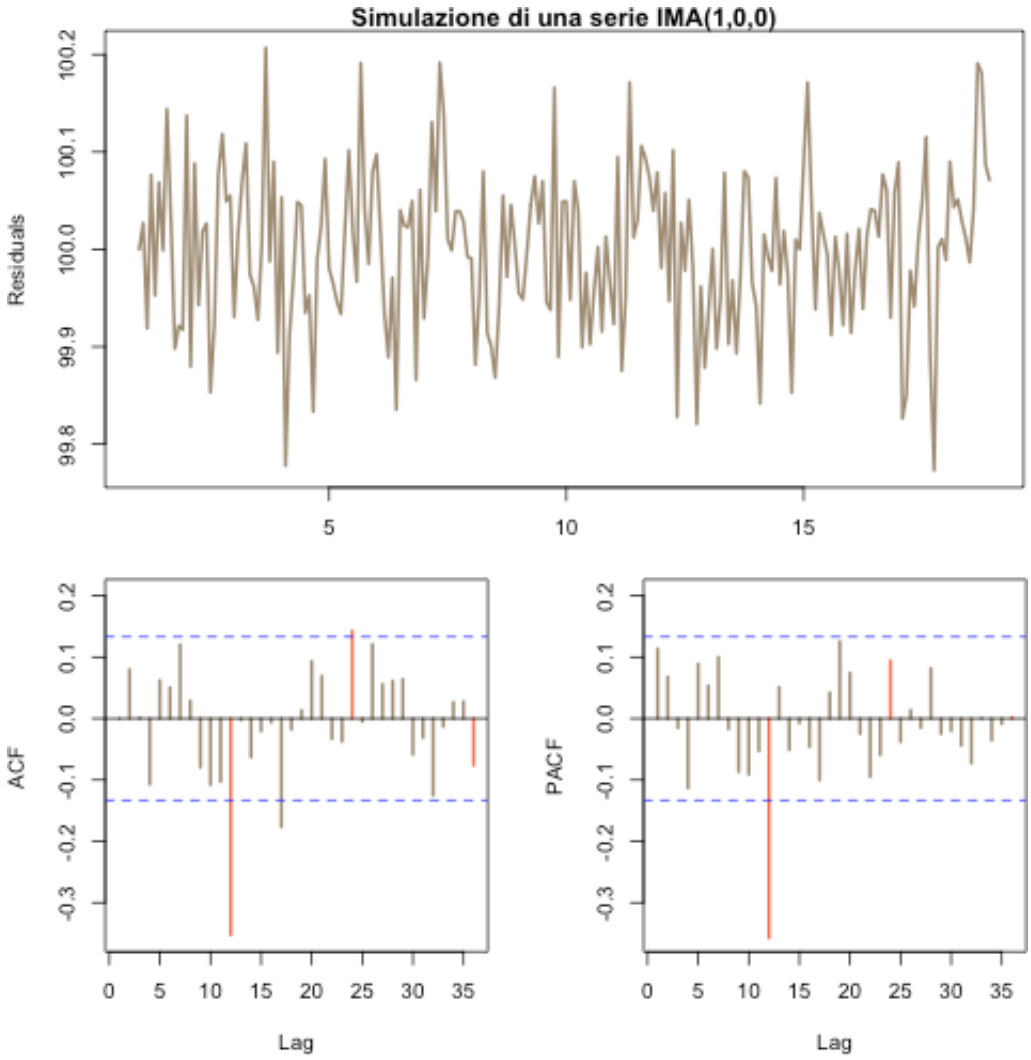


Figura 4.7: Esempio di processo con schema $\text{ARI}(1,0,0)_{12}$

L'impiego di questo modello è però controverso. La trasformata di Fourier di questo processo (ovvero lo spettro del processo) mostra picchi di uguale altezza nelle posizioni multiple della periodicità s. Non è così nella serie storica che da esso è generata, almeno per numerosità più ridotte come si vede dalla figura 4.8 in cui è riportata la rappresentazione spettrale di una serie storica $\text{ARI}(1,0,0)_{12}$. Ricordiamo però che l'uso dello spettro può scoprire solo pochi cicli ricorrenti di lunghezza differente eventualmente presenti nella serie storica.

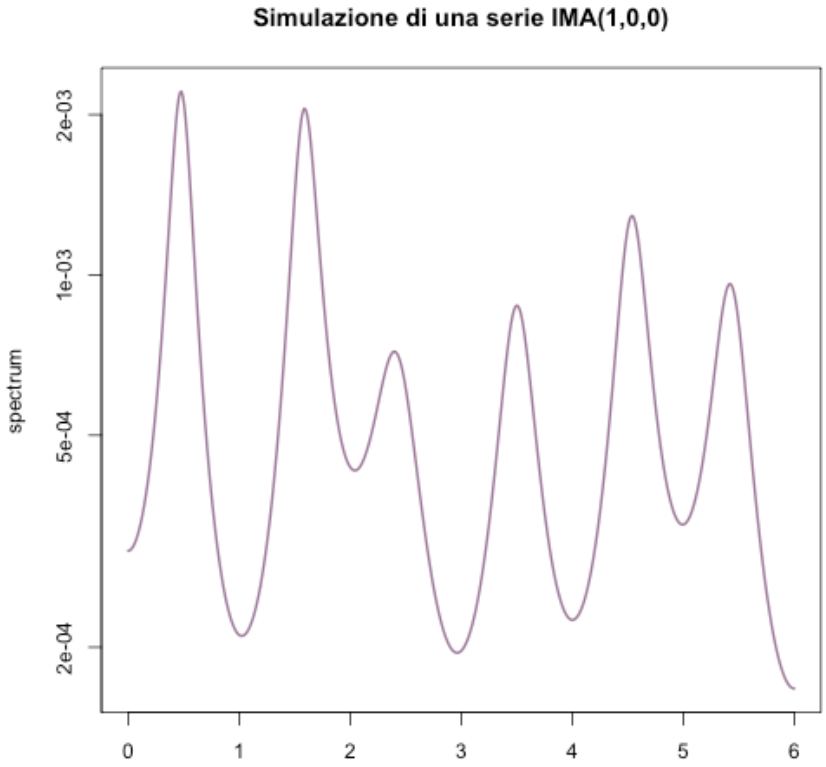


Figura 4.8: Esempio di spettro empirico di un $\text{ARI}(1,0,0)_{12}$

La teoria è corretta, ma C.W. J. Granger sostiene che questo modello possa rappresentare ogni andamento stagionale se i valori della prima annualità da cui si dipartono le varie stagionalità sono scelti in modo appropriato.

La combinazione dei due schemi più semplici porta alla formazione del modello $\text{ARIMA}(1,0,1)_s$. Per $s = 12$, l'equazione del processo è:

$$W_t = \Phi_1 W_{t-12} + a_t - \Theta_1 a_{t-12} \tag{4.19}$$

la cui funzione di autocorrelazione è data da

$$\rho(\pm sk) = \begin{cases} \dfrac{(1-\Phi_1\Theta_1)(\Phi_1-\Theta_1)}{1+\Theta_1^2-2\Phi_1\Theta_1} & \text{per } k=1 \\ \Phi_1^k & \text{per } k=2,3,\cdots \\ 0 & \text{altrimenti} \end{cases} \tag{4.20}$$

L'andamento delle **acf** e delle **pacf** per l'ARMA stagionale si comporta come quello non stagionale: decadimento progressivo per entrambi i tipi di correlazione solo che ora deve essere applicato alle sottoserie formate dai valori occupanti le posizioni dell'ordinamento temporale ricadenti nei multipli della periodicità come è ben illustrato dalla figura 4.9

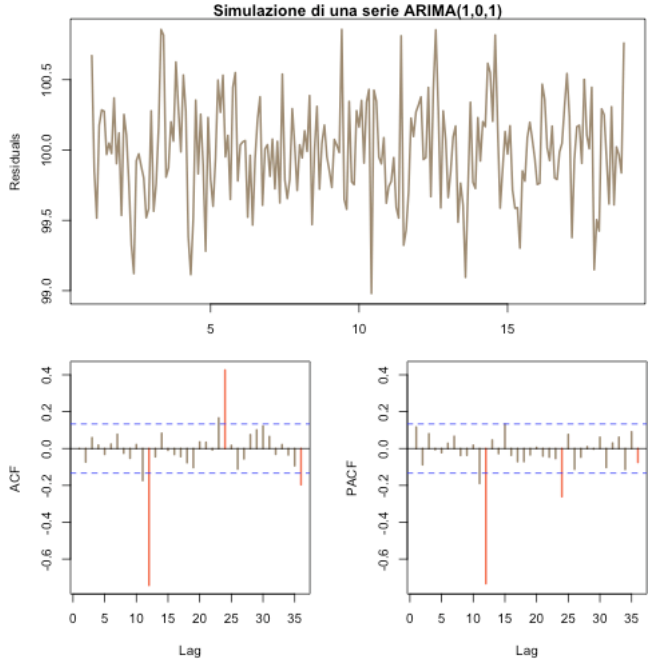


Figura 4.9: Esempio simulato di processo con schema ARIMA$(1,0,1)_{12}$

E' importante osservare che lo spettro empirico in figura 4.10 che abbiamo simulato ha una caratteristica purtroppo ricorrente. L'evidenza della stagionalità diminuisce con il progredire della serie. Questo è in contraddizione con il nostro assunto che la stagionalità sia un fenomeno connaturato alla serie e che solo in presenza di modifiche strutturali si sia autorizzati a pensare ad una modifica o ad un esaurimento della stagionalità. Poiché nulla di tutto questo è presente nella nostra tecnica di simulazione, l'apparente attenuarsi dell'effetto stagionalità è da cercarsi all'interno della formulazione ARIMA$(1,0,1)_{12}$. E' qui infatti il punto debole: la componente AR nei modelli stagionali è spesso una presenza conflittuale.

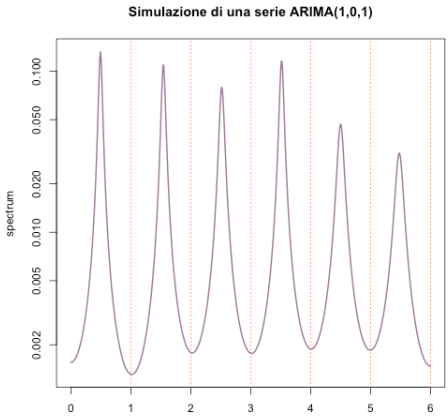


Figura 4.10: Esempio di spettro empirico di un ARIMA$(1,0,1)_{12}$

4.2.3 Modelli moltiplicativi ARIMA e SARIMA

Il ragionamento introduttivo fatto all'inizio di questo capitolo ha chiarito a sufficienza come la serie storica sia un intreccio di componenti diverse. Ognuna è pensata come risultante dalla interazione di vari fattori, non direttamente osservabili, ma ben definiti sul piano concettuale. E' fuori discussione che i fatti cambiano (e magari cambia anche il modo di definirli e/o di misurarli). Per certi elementi la variabilità è caotica e oscillatoria. Per altri è orientata in una direzione precisa ed è cumulativa. Alcuni fenomeni si modificano rapidamente, certi altri lentamente o molto lentamente. A volte l'evoluzione è così lenta da poter essere considerati fermi (almeno per un certo arco di tempo).

Il fenomeno descritto dalla serie storica diviene conoscibile attraverso le accelerazioni impresse sul movimento di fondo (che, ai fini pratici consideriamo fermo anche se non è così: ci preoccupiamo della curvatura terrestre solo per distanze che difficilmente vorremmo percorrere a piedi. Consideriamo la figura 4.11. La ragazza è ferma rispetto al cavallo che percepisce la sua allenatrice costantemente al suo fianco, ma entrambi sono in movimento. La loro velocità relativa è nulla. Solo se uno dei due accelera, l'altro lo avverte.



Figura 4.11: Accostamento di due componenti separate

Le parti della serie storica che possiamo studiare sono da considerarsi accelerazioni di diversa intensità. L'approccio classico propone di classificare i fattori che formano la parte sistematica secondo la velocità di cambiamento e di studiarli prima separatamente e poi nella loro interazione. Ogni componente agisce in modo autonomo rispetto alle altre. Tante solitudini che viaggiano su strade diverse, ma che arrivano insieme. D'altra parte, la scomposizione è necessaria dato che non siamo in grado di cogliere olisticamente la complessità dei fenomeni che producono la serie storica.

Combinazione additiva di processi regolari e stagionali

L'approccio BJ invece ammette la interdipendenza delle componenti: non stagionale e stagionale (o più componenti stagionali a diversa periodicità) e tenta di combinarle in modo credibile. Supponiamo che la parte regolare (cioè non stagionale) di una serie storica sia rappresentabile dal modello ARIMA(p, d, q)

$$\phi_p(B)(1-B)^d W_t = \mu + \theta_q(B)\, a_t \tag{4.21}$$

Ipotizziamo inoltre che la stagionalità sia riconducibile ad un SARIMA$(P, D, Q)_s$

$$\Phi_P(B^s)(1-B^s)^d W_t = \Theta_Q(B^s)\, a_t \tag{4.22}$$

Da quest'ultima abbiamo escluso il termine costante in quanto accomunato al termine costante della parte non stagionale.

Il modello SARIMA potrebbe essere legittimante interpretato come un modello ARIMA non stagionale in cui però la grande maggioranza dei parametri siano pari a zero. Ad esempio, il processo SARIMA$(1,1,2)_4$ si sviluppa come

$$\begin{aligned} \left(1-\Phi_1 B^4\right)\left(1-B^4\right) W_t &= \left(1-\Theta_1 B^4-\Theta_2 B^8\right) a_t \\ \left[1-(1+\Phi_1)B^4+\Phi_1 B^8\right] W_t &= \left(1-\Theta_1 B^4-\Theta_2 B^8\right) a_t \end{aligned} \tag{4.23}$$

e quindi

$$W_t = (1+\Phi_1) W_{t-4} - \Phi_1 W_{t-8} + a_t - \Theta_1 a_{t-4} - \Theta_2 a_{t-8} \tag{4.24}$$

che diventa un modello ARIMA$(8,0,8)$ con molti dei parametri pari a zero, ma con una evidente caratteristica di non stazionarietà poiché il coefficiente assegnato al primo termine W_{t-4} è superiore ad uno (la non stazionarietà, in questo caso deriva dalla presenza della differenza regolare di *lag* 1).

Alla stagionalità affianchiamo un modello del tipo ARIMA$(3,1,1)$ che descrive la parte non stagionale del fenomeno

$$\begin{aligned} \left(1-\phi_1 B-\phi_2 B^2-\phi_3 B^3\right)(1-B) W_t &= (1-\theta_1 B)\, a_t \\ \left[1-(1+\phi_1)B-(\phi_2-\phi_1)B^2-(\phi_3-\phi_2)B^3+\phi_3 B^4\right] W_t &= (1-\theta_1 B)\, a_t \end{aligned}$$

che comporta

$$W_t = (1+\phi_1) W_{t-1} + (\phi_2-\phi_1) W_{t-2} + (\phi_3-\phi_2) W_{t-3} - \phi_3 W_{t-4} + a_t - \theta_1 a_{t-1} \tag{4.25}$$

La somma, membro a membro, delle due equazioni (e divisione per due) consente di arrivare alla relazione

$$\begin{aligned} W_t = \frac{(1+\phi_1)}{2} W_{t-1} + \frac{(\phi_2-\phi_1)}{2} W_{t-2} + \frac{(\phi_3-\phi_2)}{2} W_{t-3} + 0.5\left[(1+\phi_1)-\phi_3\right] W_{t-4} + \\ - 0.5\Phi_1 W_{t-8} + a_t - 0.5\theta_1 a_{t-1} - 0.5\Theta_1 a_{t-4} - 0.5\Theta_2 a_{t-8} \end{aligned} \tag{4.26}$$

Siamo ancora di fronte ad un ARIMA$(8,0,8)$, ma con 8 parametri ricostruiti in base ai 7 coefficienti presenti nella modellazione della parte regolare ovvero la più lenta. La formulazione additiva risulta però inefficiente in quanto non riesce a dare conto di altri aspetti della interazione tra la componente rapida e quella lenta.

Combinazione moltiplicativa di processi regolari e stagionali

Una formulazione alternativa combina i modelli ARIMA e SARIMA in modo moltiplicativo

$$\phi_p(B)(1-B)^d(1-B^s)^D\Phi_P(B^s)W_t = \mu + \theta_q(B)\Theta_Q(B^s)a_t \tag{4.27}$$

che, nell'esempio proposto, può essere sviluppato come un ARIMA$(12,0,8)$ regolare cioè senza esplicita componente stagionale.

$$\begin{aligned}\{1-(1+\phi_1)B-(\phi_2-\phi_1)B^2-(\phi_3-\phi_2)B^3+[\phi_3-(1+\Phi_1)]B^4+\\ +(1+\phi 1)(1+\Phi_1)B^5+(\phi_2-\phi_1)(1+\Phi_1)B^6+-(\phi_3-\phi_2)(1+\Phi_1)B^7+\\ +[\Phi_1-\phi_3(1+\Phi_1)]B^8-\Phi_1(1+\phi_1)B^9-\Phi_1(\phi_2-\phi_1)B^{10}-\Phi_1(\phi_3-\phi_2)B^{11}+\\ +\Phi_1\phi_3B^{12}\}W_t = \left(1-\theta_1B-\Theta_1B^4+\theta_1\Theta_1B^5+\theta_1\Theta_2B^6-\Theta_2B^8\right)a_t\end{aligned} \tag{4.28}$$

Si può constatare che, con lo stesso numero di parametri iniziali (7) si riesce a dare conto di altri 10 coefficienti e di descrivere relazioni lineari che arrivano fino a quattro ritardi in più -nella parte autoregressiva- rispetto alla formulazione additiva. Certo, il problema è più complesso dal punto di vista computistico, ma questo non è un problema insormontabile. Inoltre, la questione della stazionarietà e della invertibilità può essere agevolmente risolta all'interno dei modelli a fattore.

Ai fini della efficacia dell'analisi è necessario ipotizzare che i polinomi $\phi_p(B)$, $\Phi_P(B^s)$,$\theta_q(B)$, $\Theta_Q(B^s)$ abbiano radici al di fuori del cerchio unitario e che quindi la relazione

$$W_t = (1-B)(1-B^s)a_t \tag{4.29}$$

rientri in un processo stocastico stazionario ed invertibile. A questo si applicano i requisiti della metodologi BJ: parsimonia (minimo numero di parametri incogniti) e previsioni solo a breve termine.

Un modello stagionale moltiplicativo molto noto per le sue applicazioni nell'ambito del traffico aereo è l'ARIMA$(0,1,1)\times(0,1,1)_{12}$ che ha equazione

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_1 a_{t-12} - \Theta_1 a_{t-13} \tag{4.30}$$

In cui Z_t corrisponde alla W_t differenziata una volta sia come componente regolare ed una volta come componente stagionale. Cleveland e Tiao (1976) hanno dimostrato che la famosa sebbene un po' obsoleta tecnica Census X_{11} usata a suo tempo dall'Ufficio centrale di statistica degli USA è molto simile al modello (4.30).

La funzione di autocorrelazione del modello "airline" è la seguente

$$\rho(\pm k) = \begin{cases} 1 & \text{per } k = 0 \\ \dfrac{-\theta_1}{(1+\theta_1^2)} & \text{per } k = 1 \\ \dfrac{\theta_1\Theta_1}{(1+\theta_1^2)(1+\Theta_1^2)} & \text{per } k = 11 \\ \dfrac{-\Theta_1}{(1+\Theta_1^2)} & \text{per } k = 12 \\ \dfrac{\theta_1\Theta_1}{(1+\theta_1^2)(1+\Theta_1^2)} & \text{per } k = 13 \\ 0 & \text{altrimenti} \end{cases} \tag{4.31}$$

Da notare che $\rho_{11} = \rho_{13} \neq 0$ il che può essere considerato un effetto della interazione (lineare) che è attiva tra la componente lenta o regolare e la componente rapida o stagionale. Inoltre, il parametro della parte stagionale non è coinvolto nella correlazione della parte non stagionale poiché ρ_1 non dipende da Θ_1; allo stesso modo la correlazione di *lag* 12 è immune dal parametro θ_1 specifico della parte non stagionale. Non è sicuro che questo sia un merito, ma ha il vantaggio di distribuire gli effetti "regular" e "seasonal" su strutture distinte.

Una formulazione additiva cosa avrebbe potuto aggiungere? La sua equazione sarebbe stata

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_1 a_{t-12} \tag{4.32}$$

in cui scompare l'addendo di *lag* 13, frutto della interazione tra componente stagionale e non stagionale che però è esclusa da questa formulazione. La funzione di autocorrelazione del modello (4.32) è la seguente:

$$\rho(\pm k) = \begin{cases} \dfrac{-\theta_1}{(1+\theta_1^2+\Theta_1^2)} & \text{per } k = 1 \\ \dfrac{\theta_1\Theta_1}{(1+\theta_1^2+\Theta_1^2)} & \text{per } k = 11 \\ \dfrac{-\Theta_1}{(1+\theta_1^2+\Theta_1^2)} & \text{per } k = 12 \\ 0 & \text{altrimenti} \end{cases} \tag{4.33}$$

Si può notare che la separazione tra la componete regolare e quella periodica è solo apparente nel modello additivo; infatti, le autocorrelazione di ritardo 1 e di ritardo 12 dipendono dai parametri di entrambe le componenti. L'interazione si esplica

per il ritardo 11, ma non per il ritardo 13 e questa è una incongruenza. Peraltro, la formulazione moltiplicativa riesce a coprire il ritardo 13 senza aggiungere nuovi parametri da stimare: in questo senso si può considerare parsimoniosa rispetto allo schema additivo.

Identificabilità dei SARIMA

Tale problema si estende anche ai modelli ARIMA stagionali. In questo senso occorre considerare l'ordine delle differenze, regolari e stagionali, nonché il grado di tutti i polinomi coinvolti. Consideriamo il modello di regressione lineare multipla

$$W_t = \beta_0 + \sum_{j=1}^{s-1} \beta_j x_{j,t} + a_t \tag{4.34}$$

che ha una parte deterministica comprendente tante variabili dicotome (*dummy*) quanti sono i termini della stagioalità (meno uno a causa della presenza dell'intercetta β_0). La relazione lineare (4.34) è virtualmente indistinguibile dal modello ARIMA stagionale

$$(1 - B^s)\, W_t = (1 - 0.95B^s)\, a_t \tag{4.35}$$

Tuttavia, il modello (4.34) contiene s parametri incogniti nella componente deterministica ed uno nella parte stocastica (la varianza dei residui a_t. Nel modello (4.35) è presente solo quest'ultimo parametetro incognito anche se si perdono s informazioni all'inizio della serie. La (4.35) può anche essere formulata nel modo che segue

$$W_t = \beta_0^{(t)} + \sum_{j=1}^{s-1} \beta_j^{(t)} x_{j,t} + a_t \tag{4.36}$$

dove $\beta_j^{(t)}, j = 0, 1, \cdots, s-1$ sono dei coefficienti variabili nel tempo che però rimangono incogniti. La difficoltà di distinguere fra i due modelli (4.34) e (4.35), solo parzialmente ridotta dai test (poco potenti) sulle radici unitarie dei polinomi, è intrinseca alla procedura BJ che infatti contempla la trattazione di molte specificazioni alternative anche sfruttando le tecniche di identificazione automatica.

4.3 Esempio applicativo

Consideriamo la serie storica relativa a *Monthly motor vehicles engines - Canada*

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1976	1404.6170	1423.2980	909.3617	1296.3190	1415.6460	1220.8750	1715.0830	1765.2500	1464.3270	1705.4490	1697.4200	1561.0600
1977	1651.2000	1636.7650	1396.3530	1196.0000	1354.2690	1650.8270	1684.4420	1232.3590	1604.6980	1472.1700	1658.4070	1869.0560
1978	1661.7820	1812.1640	1427.0540	1164.5000	1538.9110	1475.0880	1876.3160	1508.0180	1542.0860	1397.4480	1325.4920	1216.7630
1979	1205.1500	734.8333	879.9836	582.4099	1963.4680	1223.9030	708.0159	494.3333	613.7778	693.2500	731.1230	774.5538
1980	367.5757	359.7879	187.0000	299.6176	604.5588	592.2754	806.4000	617.0286	505.0845	614.1389	506.0548	605.1781
1981	728.7838	558.0667	446.6842	516.6184	821.3506	936.2949	1110.9490	700.3671	461.4557	761.2250	982.8519	1078.7320
1982	1350.8070	1453.0240	956.6786	1250.6790	458.3177	573.2235	656.4535	760.1395	781.6628	1354.0700	1819.3560	1302.7820
1983	1409.1260	1527.7500	681.4719	1364.2810	1608.4720	1708.1220	1879.0440	1616.7330	1594.3670	1790.9010	1826.4730	2136.1320
1984	2120.5720	2100.9240	1467.5870	2072.5650	2016.7720	1901.9680	1930.0540	1766.0860	1797.6810	2007.0110	1930.2770	1907.5470
1985	1814.7050	1529.0940	1019.9580	1653.8230	1837.3540	2112.0520	1717.8560	1413.9070	1513.2550	1658.5000	1486.4800	1635.3030
1986	1543.7680	1652.1820	1181.4100	1281.1800	1492.7600	1496.2080	1600.0590	1274.7430	1479.5100	1920.9410	1690.9710	1658.2810
1987	1237.2020	1345.5580	843.0857	1480.3050	1792.7430	1877.2670	1624.0380	1372.0090	1866.1320	1875.0090	2016.3550	1746.2340
1988	1914.7780	1864.4720	945.1835	1556.2840	1686.5600	1861.6730	2133.8090	1726.8090	2140.4460	1887.3780	2010.4640	1890.1160
1989	1911.9910	1681.9390	904.5652	1683.0780	1364.3300	1466.3300	1596.0600	1111.3450	1288.7690	1300.3330	1791.0420	1592.1950
1990	2022.7460	1760.0760	941.1345	1591.7900	1316.7500	1364.1320	1039.3060	597.6611	494.5360	1085.6720	764.4720	1014.9680
1991	1128.2620	1132.8570	741.2540	996.8730	1114.0320	1113.1510	979.3858	781.2460				

Figura 4.12: Monthly motor vehicles engines - Canada: 1976-1991

La serie presenta un profilo abbastanza complesso in cui appare un movimento ciclico sia pure irregolare che potrebbe essere rappresentato da un polinomio di grado sette dato che tanti sono i picchi e le valli armonicamente presenti. Tuttavia, come si vede dal grafico, gran parte dei valori della serie rimarrebbero tagliati fuori e forse si perderebbe la parte più interessante costituita dai rialzi e dai ribassi dovuti ad effetti diversi da quelli provocati dal trend.

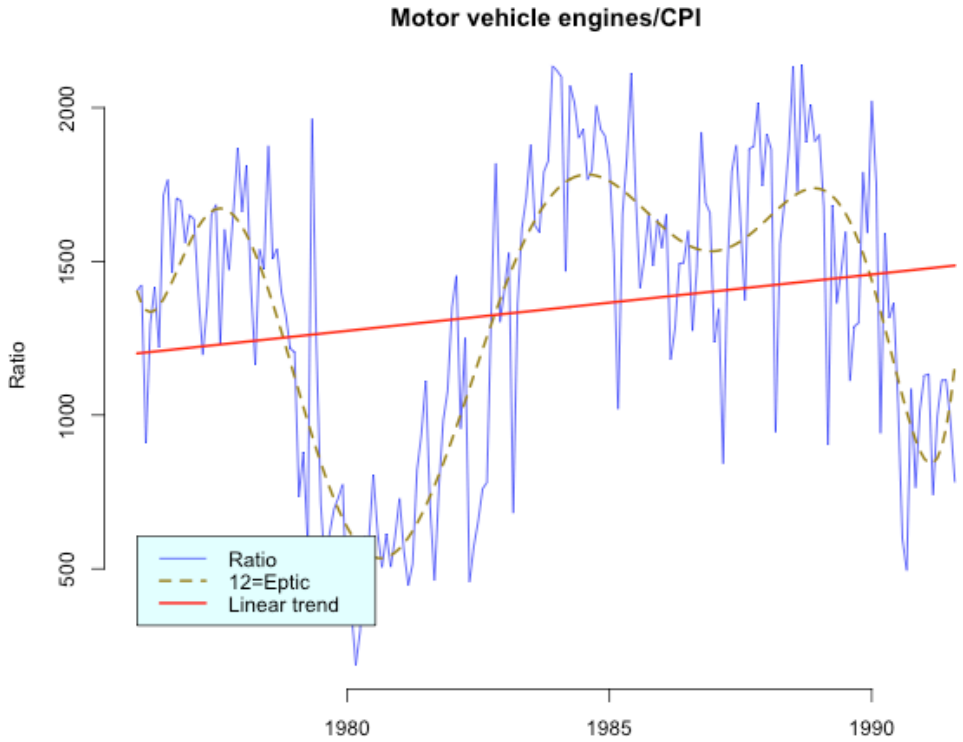


Figura 4.13: Monthly motor vehicles engines - Canada. Analisi del trend

La presenza della stagionalità non è pacifica. Infatti, l'analisi della varianza effettuata sulla serie storica depurata dal trend produce un valor p intorno a $4x10^{-8}$. Anche l'andamento dello spettro relativo alla serie storica in studio conferma la presenza di una discreta stagionalità che però non è del tutto stabile e richiederebbe un modello di stagionalità variabile se ci trovassimo nel contesto dell'approccio classico. Infatti, le frequenze stagionali si intensificano e si attenuano nell'evolversi del fenomeno con dei periodi in cui l'effetto stagionale è assente ovvero oscurato dall'impatto di altri fattori.

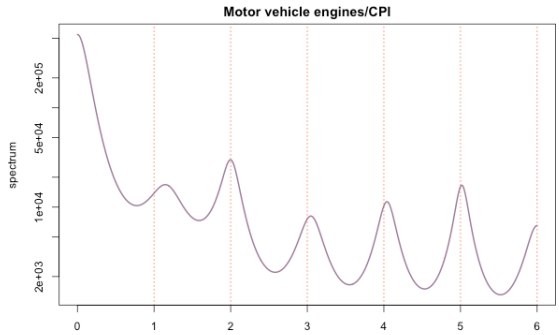


Figura 4.14: Analisi spettrale per lo studio della stagionalità

L'andamento ciclico dei *motor vehicles engines* non prelude ad un grande effetto per le differenziazioni dato che il polinomio che la rappresenta ha un grado troppo elevato. Tuttavia, nella figura 4.12 si intravvede l'effetto di una tendenza lineare che non deve essere trascurata. Nella figura 4.15 si può osservare l'influenza stabilizzante dovuta alle differenze prime (da notare anche il rimbalzo attribuibile all'effetto del valore anomalo prima evidenziato). E' poi sicuro che l'ordine delle differenze non deve essere superiore a due dato che le terze e le quarte differenze inducono un aumento devastante della entropia nella serie. In definitiva, nella ricerca dell'ordine delle differenze per la componente "regular porremo il termine a due proprio per non escludere le possibilità che abbiano anche solo la parvenza della plausibilità.

Rimane da accertare se non sia il caso di procedere alla differenziazione stagionale. Nella figura 4.15 si può notare che le l'operatore $(1 - B^{12})$ non produce alcuna modifica sostanziale al grafico della serie ricopiandone la struttura. L'operatore $(1 - B^{12})^2$ ottiene un risultato peggiore, ma per certi versi più interessante: rende uniforme l'ordine della variabilità ... peggiorandolo, ma senza scompensi nei vari tratti dell'arco temporale di osservazione; inoltre, cancella, o forse è meglio dire, oscura l'azione del valore anomalo che precede l'anno 1980. Poiché la serie mensile in studio non è molto distesa contando su $n = 188$ osservazioni, non ha molto senso coinvolgere le differenze stagionali di ordine due, ma l'effetto sul valore anomalo

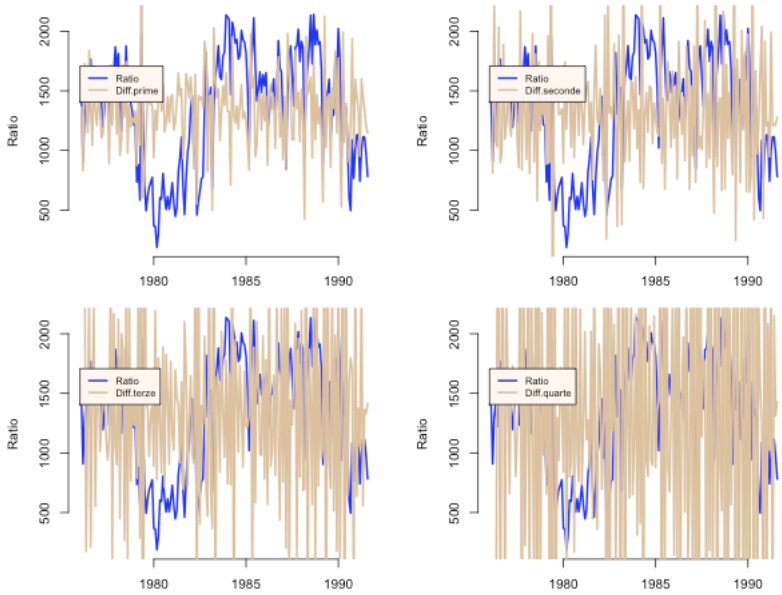


Figura 4.15: Valutazione dell'ordine delle differenze regolari

rimane illuminante. Si può poi aggiungere che nella zona centrale dell'andamento della serie c'è un comportamento dissimile dal resto che non può essere attribuito al ruolo, sia pure rilevante, di un singolo *outlier*. E' chiaro che se la modellistica pura ARIMA non funziona bene, dovremo ricorrere ad una tecnica in cui il ciclo-trend sarà stimato a parte e la tecnica BJ applicata ai suoi residui.

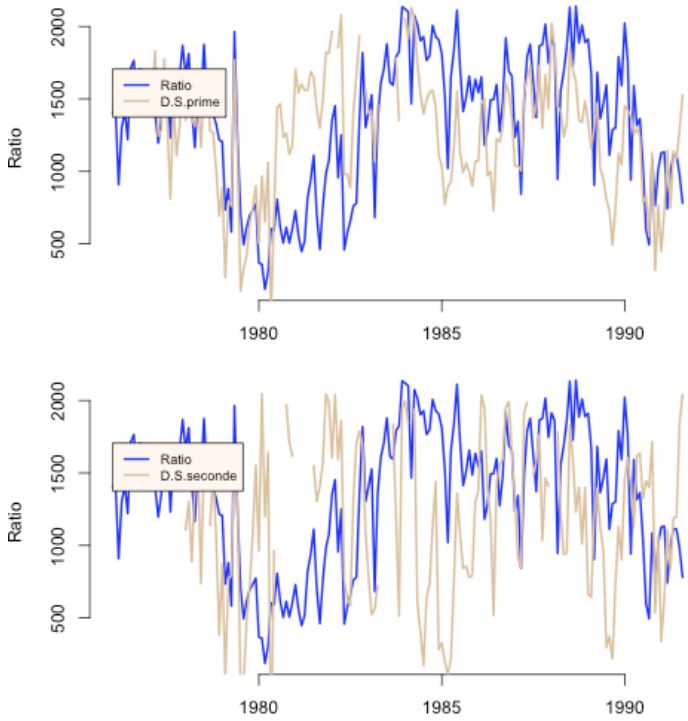


Figura 4.16: Valutazione dell'ordine delle differenze stagionali

4.3.1 Studio dei correlogrammi

Nella figura 4.17 è riportato il correlogramma della differenze. In evidenza c'è la componente autoregressiva che è presente sia nella parte regolare che in quella stagionale. Il grafico delle auto-correlazioni parziali sembra preludre ad una componente media mobile nella parte regolare forse di ordine uno e forse di ordine due, non è chiaro. Anche la parte stagionale ha una chiara struttura media mobile con un ordine che solo per motivi di dimensione della serie limiteremo ad uno. La figura 4.17 segnala come auto-correlazioni parziali significative quella al lag 1 e quella al lag 4, ma per valori negativi segno che una la parte regolare potrebbe avere un struttura auto-correlativa almeno fino al quarto ritardo ovvero una struttura ARMA molto articolata. In definitiva, l'ambito della scelta per le procedure di selezione automatica dovrebbe essere

$$1 \leq p \leq 4; \qquad 0 \leq d \leq 1; \qquad 0 \leq q \leq 2$$
$$0 \leq P \leq 2; \qquad 0 \leq D \leq 1; \qquad 0 \leq Q \leq 1$$

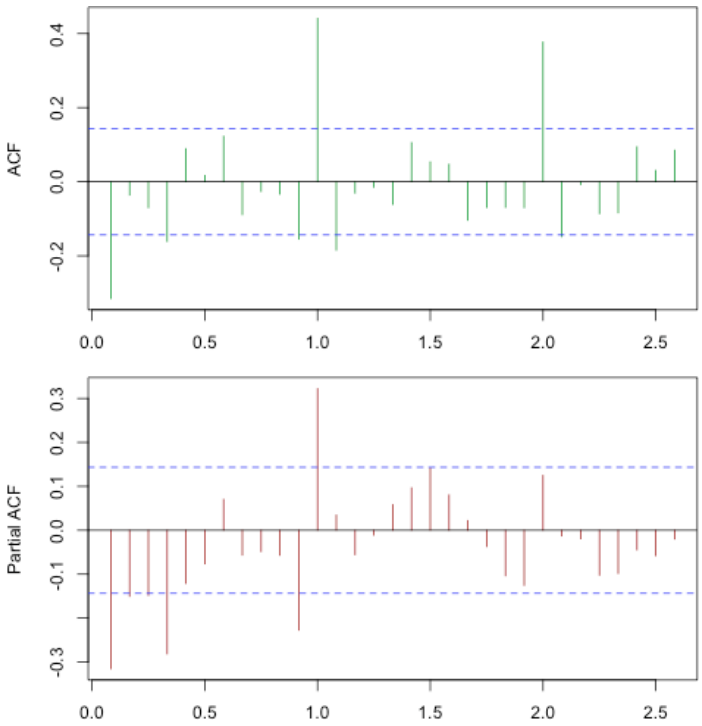


Figura 4.17: Correlogramma della serie *motor vehicles engines*

CAPITOLO 5

Stima, Identificazione e validazione degli ARIMA

5.1 Stima

Consideriamo il processo ARMA stazionario

$$\Phi_P(B^s)\,\phi_p(B)\,W_t = \mu + \theta_q(B)\,\Theta_Q(B^s)\,a_t \tag{5.1}$$

con a_t *white noise* gaussiano. La parte autoregressiva -regolare e stagionale- è descritta, rispettivamente, dai polinomi $\phi_p(B)$ e $\Phi_P(B^s)$; la parte media mobile è descritta dai polinomi $\theta_q(B)$ e $\Theta_Q(B^s)$. Tenuto conto dell'ordine dei polinomi, la determinazione del processo ARMA in (5.1) pone il problema di stimare un numero di parametri incogniti pari a

$$P + p + Q + q + 1 = m \tag{5.2}$$

E' vero che il loro numero potrebbe essere contratto vincolando al valore zero quelli che sono a priori nulli oppure che lo risultano dopo un'analisi preliminare, ma potrebbero comunque essere ancora tanti. Inoltre, la stima con vincoli è ottima fintanto che i vincoli stessi sono rispettati in modo naturale dal modello, ma l'ottimo scompare non appena i vincoli cominciano ad imporsi. Infatti non c'è alcuna garanzia che l'ottimo vincolato coincida con l'ottimo globale.

5.1.1 Stima di massima verosimiglianza

La procedura di stima dovrebbe seguire il principio analogico: si sceglie lo stimatore che nel campione risolve il problema risolto nella popolazione dalla sua controparte teorica (anche grazie al supporto fornito dalla teoria asintotica che tende in genere ad avvalorare questa scelta). Si consideri un insieme di n valori ottenuto con lo schema

del campionamento casuale semplice, cioè osservazioni indipendenti ed identicamente distribuite

$$S = \{w_1, w_2, \cdots, w_t, , \cdots, w_n\} \tag{5.3}$$

Sia $F(W, \theta)$ la funzione di ripartizione che governa la casualità delle estrazioni con $\boldsymbol{\eta}$ vettore di parametri incogniti da stimare. Qual'è il valore di $\boldsymbol{\eta}$ che più verosimilmente (cioè con maggiore probabilità) ha indotto la sequenza a produrre proprio i valori che sono stati osservati?

Per ogni valore di $\boldsymbol{\eta}$ la probabilità che si realizzi S è data da

$$Pr[S|\boldsymbol{\eta}] = \Pi_{t=1}^{n} f(W_t, \boldsymbol{\eta}) \tag{5.4}$$

dove $f(W_t, \boldsymbol{\eta})$ è la funzione di densità della variabile casuale W_t che, per semplicità di ragionamento, ipotizziamo continua. Se la densità di probabilità per valori vicino a w_t è piccola, si può ritenere improbabile che si realizzino valori prossimi a w_i; se è grande, allora i valori vicini a w_t potranno essere più frequenti. La probabilità congiunta dei valori osservati è proprio la funzione di verosimiglianza allorché la si considera come una funzione dei parametri incogniti $\boldsymbol{\eta}$

$$L(\boldsymbol{\eta}|w_1, w_2, \cdots, w_n) = \sum_{t=1}^{n} Log[f(W_t, \boldsymbol{\eta})] \tag{5.5}$$

in cui la trasformata logaritmica semplifica, ma non modifica il comportamento della funzione di verosimiglianza. Lo scopo della stima ML è di cercare il valore di $\boldsymbol{\eta}$ che renda massima la probabilità di osservare i valori riscontrati, cioè la serie storica S.

Esempio 9. *Il modello gaussiano è tra i più usati in statistica sia perché sorretto dal teorema limite centrale e sia (forse soprattutto) perché si accompagna a soluzioni efficienti ed immediate anche se non necessariamente appropriate, grazie al criterio dei minimi quadrati. Supponiamo che ciascuna delle variabili casuali della serie storica abbia densità gaussiana* $N(\mu, \sigma_w^2)$

$$f(W_t, \boldsymbol{\theta}) = \frac{e^{-0.5\left(\frac{w_t - \mu}{\sigma_w^2}\right)^2}}{\sqrt{2\pi\sigma_w^2}}$$

La corrispondente funzione di verosimiglianza è data da

$$L\left(\mu, \sigma_w^2|w_1, w_2, \cdots, w_n\right) = -\frac{n}{2} Log\left(2\pi\sigma_W^2\right) - \frac{1}{\sigma_w^2} \sum_{t=1}^{n} Log[f(W_t, \boldsymbol{\eta})]$$

Poiché la $L(\cdot)$ *è stata ipotizzata differenziabile rispetto ai parametri su tutto il loro dominio, il massimo si può determinare con le derivate parziali. In particolare*

$$\frac{\partial L(\cdot)}{\partial \mu} = -\frac{1}{2\sigma_w^2} \sum_{t=1}^{n} (w_t - \mu) 2(-1) = \frac{1}{\sigma_w^2} \sum_{t=1}^{n} (w_t - \mu) = 0$$

e quindi la stima ML di μ è la media aritmetica semplice dei valori osservati. Per il secondo parametro si ha

$$\begin{aligned}\frac{\partial L(\cdot)}{\partial \sigma_W^2} &= -\frac{n}{\sigma_w^2} - \frac{1}{2}\left[\sum_{t=1}^{n}(w_t-\mu)^2\right]\frac{-2}{\sigma_W^3} = 0 \\ &= \frac{\left[\sum_{t=1}^{n}(w_i-\mu)^2\right]}{\sigma_W^2} - n = 0 \rightarrow \sigma_W^2 = \frac{\left[\sum_{t=1}^{n}(w_t-\mu)^2\right]}{n}\end{aligned}$$

Pertanto, la stima di σ è lo scarto quadratico medio dei valori osservati.

La soluzione è semplice, riconoscibile ed elegante dal punto di vista formale e con calcoli puttosto semplici. Tutte queste proprietà derivano dall'ipotesi di gaussianità delle distribuzioni marginali. A queste si aggiungono le caratteristihe asintotiche e campionarie tipiche degli stimatori di massima verosimiglianza che si fondono con quelle dei minimi quadrati.

Se conoscessimo la funzione di densità $f(W, \boldsymbol{\eta})$ potremmo ripetere lo studio di verosimiglianza anche per altre distribuzioni, ma questo non è quasi mai vero. Potremmo forse avere qualche indicazione sulla asimmetria, sulla unimodalità, sull'eventuale grado di appuntimento al centro e sullo spessore delle code, ma molto difficilmente questi indizi si trasformano nella specificazione di un modello riconoscibile di variabile casuale.

5.1.2 Stima di massima verosimiglianza degli ARIMA

Partiamo dalla serie storica $W_t, t = 1, 2, \ldots, n$; quindi con dati disponibili per n periodi ed ipotizziamo perciò di avere già effettuato l'imputazione dei valori mancanti ed il trattamento dei valori remoti, oltre le necessarie correzioni di calendario e feste mobili. Supponiamo inoltre che la distribuzione marginale del processo sia gaussiana e che sia nota la varianza dell'associato *white noise* con σ_a^2. Riscriviamo il modello (5.1) eplicitando meglio il ruolo del *white noise*.

$$a_t = \left[\frac{\Phi_P(B^s)\,\phi_p(B)}{\theta_q(B)\,\Theta_Q(B^s)}\right](W_t - \mu) \tag{5.6}$$

Il logaritmo della funzione di verosimiglianza del processo, dal punto di vista del *white noise* può essere espresso con

$$L(\boldsymbol{\Phi}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\Theta}, \mu | a_1, a_2, \cdots, a_n) = -\frac{n}{2} Log\left(2\pi\sigma_a^2\right) - \frac{1}{\sigma_a^2}\sum_{t=1}^{n} a_t^2 \tag{5.7}$$

In questo scenario, i parametri incogniti compaiono solo nel secondo addendo per cui la loro stima deve essere rivolta alla minimizzazione (a causa del segno negativo)

della sommatoria con il metodo dei minimi quadrati.

$$S\left(\mathbf{\Phi}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mathbf{\Theta}, \mu\right) = \sum_{t=1}^{n} a_t^2 \tag{5.8}$$

Non si conosce nulla della serie storica degli a_t che, peraltro, deve essere implicitamente considerata estesa anche a dei valori precedenti l'inizio del processo. Ad esempio, per un ARMA all'epoca t dovremmo avere conoscenza dei *white noise* passati $\tilde{\mathbf{a}} = (\tilde{a}_{t-1}, \tilde{a}_{t-2}, \cdots, \tilde{a}_{t-q-Q_s})$ dove s è la stagionalità. Per il momento è comodo ipotizzare che questi errori precedenti l'inizio della serie storica siano pari a zero che è poi, in fondo, il loro valore atteso: $E\left(\tilde{a}_t\right) = 0 \; \forall \; t$.

Se conoscessimo il *white noise* al ritardo a_0 potremmo risalire gradualmente agli altri. Analizziamo ad esempio un ARMA$(0, 0, 1)$ con

$$\begin{aligned} W_t &= a_t - \theta_1 a_{t-1} \\ W_t + \theta_1 a_{t-1} &= a_t \longrightarrow a_t = W_t + \theta_1 a_{t-1} \end{aligned}$$

Noto a_0, arriviamo ad a_1 con $a_1 = W_1 + \theta_1 a_0$ dato che W_1 è un valore che è stato effettivamente osservato. Analogamente possiamo determinare a_2 conoscendo a_1

$$a_2 = W_2 + \theta_1 a_1 = W_2 + \theta_1 \left(W_1 + \theta_1 a_0\right) = W_2 + \theta_1 W_1 + \theta_1^2 a_0$$

Anche W_2 è stato osservato e quindi si può calcolare a_2 e così via fino ad arrivare ad a_n che richiede la conoscenza sia di W_n che di a_{n-1}. Quindi, la conoscenza del *white noise* iniziale a_0 ci permette di ricostruire gli altri termini di errore. Se lo ipotizziamo nullo come si è illustrato in precedenza, possiamo effettuare la massimizzazione della verosimiglianza condizionata da $a_0 = 0$. Per il calcolo della somma in (5.8) dovremmo però conoscere anche alcuni dei valori del processo $\tilde{W}_t$ precedenti la prima osservazione: $\tilde{W}_{t-1}, \tilde{W}_{t-2}, \ldots, \tilde{W}_{t-p-P_s}$. A questo fine possiamo seguire diverse strategie

- Considerarli uguali alla media globale $\hat{\mu}$ dei valori osservati della serie storica e procedere con la stima di massima verosimiglianza condizionata:

  $$S\left(\mathbf{\Phi}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mathbf{\Theta}, \mu | \tilde{\mathbf{a}}\right)$$

  in cui i valori passati $\tilde{W}_{t-1}, \tilde{W}_{t-2}, \ldots, \tilde{W}_{t-p-P_s}$ sono considerati fissi. E' una ipotesi arbitraria, ma dall'effetto transeunte data la stazionarietà del processo.
- Ipotizzare l'avvio della serie storica da un ritardo successivo all'ultimo valore non osservato e utilizzare i valori risparmiati per stimare quelli mancanti.

- Effettuare i cosiddetto *backforecasting* cioè applicare la prima strategia alla serie con ordine temporale invertito cioè dal *lag* n al *lag* 1 e procedere a far scorrere in avanti il modello stimato le cui previsioni diventano stime dei valori passati per la serie nell'ordine temporale originario.

Una volta completata la somma in (5.8) si può procedere alla sua minimizzazione. Occorre subito dire che si tratta di una funzione non lineare nei parametri incogniti anche se questo, a prima vista, può sorprendere dato che i processi sono considerati lineari. La linearità però è solo apparente. Non si deve dimenticare che i processi W_t e a_t si scambiano di ruolo ad ogni periodo: incognite al tempo t e valori noti nel periodo successivo $t+1$ cosicché i parametri ad essi associati cominciano a moltiplicarsi e intrecciarsi causando non-linerità.

La stima non lineare negli ARIMA si avvale di algoritmi efficienti e consolidati quali, ad esempio, il noto schema di Levenberg-Marquardt. Tali tecniche procedono alla ricerca del minimo sfruttando il valore della funzione di verosimiglianza e le sue derivate prime e seconde (talvolta approssimate con delle differenze). Le procedure iterative, nel loro cammino, possono incontrare punti di ottimo che però corrispondono ad AR non stazionari o ad MA non invertibili. Questi due requisiti sono entrambi necessari per non incorrere in errori standard infiniti degli stimatori dei parametri. Il *software* per il trattamento delle serie storiche è in grado di aggirare questa difficoltà, ma non è sicuro che il punto di ottimo globale per $\text{Ł}(\cdot)$ rispetti le caratteristiche volute per gli ARIMA. In genere, non si ha bisogno di controllare in che modo abbia proceduto l'algoritmo di calcolo, a meno di non sospettare minimi locali.

La varianza degli a_t non è nota come si è ipotizzato all'inizio, ma deve essere stimata come tutti gli altri parametri. Una comodità offerta dal metodo dei minimi quadrati ovvero dalla gaussianità delle marginali del processo a_t è che la stima di σ_a^2 è un risultato collaterale alla stima dei parametri e può avvenire separatamente e in una fase successiva. Infatti

$$\hat{\sigma_a^2} = \frac{\sum_{t=1}^{n} \hat{a}_t}{n-m}; \quad \hat{\mu} = \frac{\sum_{t=1}^{n} W_t}{n} \tag{5.9}$$

La stima della varianza degli a_t è subito coinvolta nella matrice di varianze-covarianze dei parametri $\mathbf{C} = (c_{ij})$ in essa inclusi ed alle statistiche tipo t-Student. I parametri che risultano poco significativi andrebbero esclusi la procedura di stima ripetuta usando solo le posizioni collegate a parametri il cui valor p approssimato sia inferiore, ad esempio, all'1%.

Si dimostra che gli stimatori massima verosimiglianza dei processi ARIMA con errori gaussiani sono almeno asintoticamente corretti, consistenti, asintoticamente

gaussiani e asintoticamente efficienti. Una parte di queste proprietà deve essere accantonata se l'ipotesi di gaussianità non ha riscontro nei dati.

5.2 Adeguatezza del modello

Conoscere egregiamente i dati statistici di un problema non ci esime dal controllo dei ragionamenti di fondo che ne governano o dovrebbero governarne l'impostazione. La scelta del modello ARIMA non deve scaturire da affezione al suo semplice schematismo, ma anche alla qualità dei risultati previsionali che con esso si ottengono. Se la serie storica fosse correttamente rappresentata da un modello ARMA del tipo

$$a_t = \left[\frac{\Phi_P\left(B^s\right)\phi_p\left(B\right)}{\theta_q\left(B\right)\Theta_Q\left(B^s\right)}\right]\left(W_t - \mu\right) \tag{5.10}$$

allora i residui stimati $\hat{a}_t$ ad esso attribuibili dovrebbero risultare privi di ogni struttura di dipendenza lineare; cioè le autocorrelazioni semplici e parziali risulterebbero poco significative e la loro l'evoluzione temporale essere erratica come un *white noise* sotto tutti i punti di vista. In sintesi, si può dire che questo sia lo scopo dell'analisi delle serie storiche: filtrare la serie dei valori osservati fino a lasciare solo ciò che si può esprimere come un processo lineare stazionario semplice privo di memoria e gaussiano (ovvero un *white noise*) recuperando dal filtro le componenti strutturali e più interessanti.

5.2.1 Misure di adattamento

E' raro che una procedura di stima statistica di un qualsiasi modello si possa concludere con la convinzione che i risultati siano i migliori ottenibili dalla combinazione di dati, procedure ed algoritmi di calcolo. Ad esempio l'errore standard dei residui dà un'indicazione univoca solo nel caso sia nullo, altrimenti non si riuscirà mai a dire, senza ulteriori informazioni, se è basso oppure alto. Anche nel contesto dei modelli Box-Jenkins si rende necessario una misura della bontà di adattamento tra modello ipotizzato e dati osservati.

L'indice di adattamento più comune per verificare la validità di un modello ARIMA, regolare o stagionale, è quello di Ljung-Box

$$R_h = \sum_{i=1}^{h}\left[\frac{n\left(n+2\right)}{n-i}\right]\hat{\rho}_i^2 + \frac{h\left(h+1\right)}{2n} \tag{5.11}$$

dove h è il ritardo massimo a cui ci si aspetta di trovare autocorrelazioni significativamente diverse da zero; dopo h la stazionarietà dovrebbe rendere virtualmente

nulle tutte le autocororrelazioni. Le soglie critiche di h per decidere in un senso o in un altro, proposte in letteratura, sono più di una. Ad esempio

$$h = \min\left\{30, \frac{n-1}{6}\right\}; \quad h = 10 + \sqrt{n}; \quad h = \left\lfloor 12\left(\frac{n}{100}\right)^{0.25} + 0.5\right\rfloor \tag{5.12}$$

Se i residui del modello (5.10) si comportassero come un *white noise* allora la (5.11) dovrebbe dar luogo a valori contenuti, per ogni ragionevole valore di h.

R_h è un *test* di tipo *portmanteau* cioè un misto di più fattori: le autocorrelazioni ed i pesi decrescenti con cui queste sono ponderate. La distribuzione di riferimento è quella del χ^2 con $(h-m)$ gradi di libertà, dove $(m = p + P + q + Q + 1)$. Una valutazione grezza, ma immediata si può basare sul ritardo massimo h. Tenuto conto che in una χ^2, i gradi di libertà sono pari alla sua media, qualora risulti che $R_h > h - m$, c'è da ritenere che non tutto il contenuto informativo dei dati sia stato assorbito dal filtro ARIMA e che potrebbe esserci un modello migliore soprattutto a fini previsivi. Più formalmente, se il valor p della statistica R_h è inferiore all'1% si può rifiutare l'ipotesi che gli $\hat{a}_t$ siano privi di struttura.

In alternativa ad R_h, si può considerare l'indice proposto da [**?**]

$$R_h^* = \sum_{i=1}^{h} \left[\frac{n\,(n+2)}{n-i}\right] \hat{\phi}_i^2 \tag{5.13}$$

che è pure compatibile con un *test* χ^2 con $h - m$ gradi di libertà. Valori elevati di R_h^* indicano che si è forse omessa qualche componente ed il modello finale proposto deve essere rivisto; non sarà così se il p-value di R_h^* è superiore al 5%.

Le valutazioni riferite a fattori stocastici, come è ben noto, debbono sempre essere improntate alla prudenza. L'ipotesi sottoposta a verifica è la seguente:

$$\begin{cases} H_0: & \text{il modello esaminato è adeguato} \\ H_1: & \text{ripetere l'analisi con un altro modello} \end{cases} \tag{5.14}$$

E' bene ricordare che le distribuzioni di probabilità gaussiana e del χ^2 con le loro soglie di riferimento scaturiscono da approssimazioni valide sotto condizioni discutibili e di difficile verifica oppure inesistenti e che in ogni caso richiedono periodi di osservazioni molto lunghi. Sono accettate come atto di fede derivato dall'originale dogma che la gaussianità non produce sfracelli caso mai fosse violata. In poche parole, il valor p che si accompagna al test ha più una valenza qualitativa che quantitativa.

5.2.2 Misure di adattamento e parsimonia

La valuzione della bontà di adattamento di un modello statistico deve tenere conto del fatto da tempo conglamato che l'accordo tra valori stimati ed i corrispondenti valori osservati aumenta e talvolta sensibilmente, con l'aumentare del numero

di parametri, senza che si possa riscontrare una aumentata capacità predittiva del modello. Per arginare questa enfatizzazione impropria dell'abilità di una certa formulazione, sono stati proposti diversi criteri che hanno come denominatore comune l'essere somma di due addendi riferibili ad aspetti distinti dell'adattamento.

La prima parte è quella derivata dalla somma dei residui osservati $\hat{a}_t$ al quadrato (quasi sempre in scala logaritmica), che conclude la procedura di identificazione e stima del modello. Questa componente diminuisce con l'aumentare della bontà di adattamento. Il secondo addendo è una quantità che è invece crescente con il numero di parametri incogniti complessivamente stimati, in modo da penalizzare le scelte non parsimonionse ovvero i modelli che sono in *overfitting*. La differenza tra le molte misure escogitate a questo proposito risiede nel modo di bilanciare l'impatto dei due addendi.

$$\begin{aligned}
\text{Akaike information criterion}(AIC) &: C = \log\left(\hat{\sigma}_a^2\right) + \frac{2m}{n} \qquad (5.15)\\
\text{Hurvich-Tsay (AICC)} &: C = \log\left(\hat{\sigma}_a^2\right) + \frac{2m}{n-m}\\
\text{Hannan-Quinn (HQIC)} &: C = \log\left(\hat{\sigma}_a^2\right) + \frac{2m\log\left[\log\left(n\right)\right]}{n}\\
\text{Bayesian information criterion (BIC)} &: C = \log\left(\hat{\sigma}_a^2\right) + \frac{m\log\left(n\right)}{n}\\
\text{Shi-Tsay (RIC)} &: C = (n-m)\log\left(\hat{\sigma}_a^2\right) - m\left[\log\left(n\right) - 1\right]\\
&\quad -\frac{4}{n-m-2}
\end{aligned}$$

Non è del tutto da trascurare l'indice di adattamento più familiare nell'ambito del modello di regressione lineare cioè il coefficiente di determinazione corretto per i gradi di libertà

$$\hat{R}^2 = 1 - \left[\frac{n-1}{n-m}\right]\left(1 - \frac{\hat{\sigma}_a^2}{\sigma_W^2}\right) \qquad (5.16)$$

Se il modello ha un valore di $\hat{R}^2$ prossimo all'unità potrà considerarsi soddfisfacente, ma se è prossimo allo zero o negativo, il modello in esame è da scartare. In questo senso si può dire che le indicazioni negative sono da prendersi più sul serio di quanto non si possa fare con quelle positive.

Ognuno degli indici in (5.15) è in grado di orientare la scelta del modello più appropriato per la serie storica osservata, almeno per quel che riguarda la bontà di adattamento. Tuttavia, per asseverare la scelta di un modello ARIMA occorre considerare altri aspetti

- Significativà dei coefficienti stimati (t-Student con valor p minori dell'1%).
- Azzeramento sostanziale dei coefficienti di autocorrelazione per la serie dei residui stimati.

- Radici esterne al cerchio unitario sia per AR che per MA in modo da assicurare la stazionarietà e l'invertibilità. Se le radici sono prossime all'unità forse è bene effettuare delle differenziazioni o, addirittura, cambiare modello.
- Erraticità dei residui stimati.

5.2.3 Indici di erraticità

Nelle serie storica c'è sempre una componente casuale, stocastica o erratica che dovrebbe essere l'unica che rimane dopo che si sia trovato un buon modello. Errare ha molti significati:

1. Andare di qua e di là senza meta o senza scopo; vagare, peregrinare (anche fig.): andar vagando (senza saper dove, senza consiglio, come brancolando fra le tenebre),errare per la città, per i campi; errare con lo sguardo, andare qua e là senza una direzione certa, errare con la fantasia. Percorrere vagando: Dante... errava / pensoso peregrin la selva fiera.
2. In base alla radice gotica: scorrere, sdrucciolare
3. (lett.) Allontanarsi, deviare dal retto sentiero, fuorviarsi, deviare dal vero.
4. Cadere in errore; ingannarsi, sbagliare, fallire.

L'approccio Box-Jenkins punta sul significato di allontanamento da un sentiero predefinito che il fenomeno segue a meno di piccoli ed incontrollabili scarti imprevedibili se non nel loro comportamento generale. Come se ci fosse un punto, una stella polare che magnetizza la direzione del fenomeno oppure fornisca la bussola per orientarsi.

Per accertare la completa erraticità di quello che il modello ha lasciato nei residui si può effettuare un test di casualità o erraticità. Ne esistono diversi. Secondo Kendall dovrebbero:

(•) Fare ipotesi poco restrittive sulla distribuzione da cui si suppone provengano gli errori.

(•) L'idea ispiratrice deve essere semplice e non comportare calcoli onerosi.

(•) Lo schema di calcolo deve essere aggiornabile ad ogni nuovo dato senza dover riprendere tutti i calcoli dall'inizio.

Test dei punti di svolta

Tra i vari procedimenti di verifica della casualità quelli più spesso inseriti nel *software* delle serie storiche c'è il *test* dei punti di svolta. Si conta il numero di svolte: picchi e valli. Una svolta richiede 3 punti consecutivi: a_{t-1}, a_t, a_{t+1}.
in quattro casi c'è un punto di svolta nel centro della terna. In una serie di n elementi esistono $n-2$ terne consecutive (esclusi il primo e l'ultimo termine). Se la serie

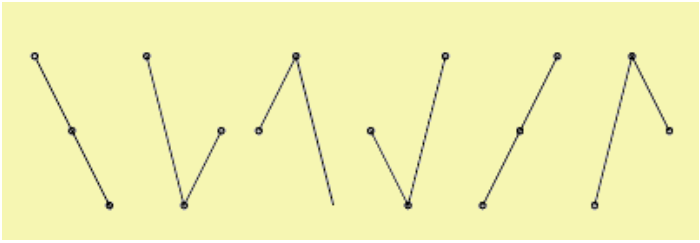


Figura 5.1: *Test* dei segni

fosse del tutto erratica, allora i $3! = 6$ possibili ordinamenti sarebbero equipresenti, almeno tendenziamente. Tra questi solo quattro contengono un punto di svolta. A partire da una variabile indicatore dei punti di svolta:

$$I_t = \begin{cases} 1: & \text{se } a_{t-1} < a_t > a_{t+1} \text{ oppure } a_{t-1} > a_t < a_{t+1} \\ 0: & \text{altrimenti} \end{cases} \tag{5.17}$$

In caso di valori coincidenti per i residui (un fatto un po' inverosimile per variabili metriche) il valore degli errori si considera unico riducendo n. Il numero totale dei punti di svolta è

$$\nu = \sum_{t=2}^{n-2} I_t \tag{5.18}$$

La statistica ν ha valore atteso e varianza data da

$$E(\nu) = \sum_{t=2}^{n-2} E(I_t) = \sum_{t=2}^{n-2} \frac{4}{6} = \frac{2(n-2)}{3} \tag{5.19}$$

$$Var(\nu) = \frac{16n - 29}{90} \tag{5.20}$$

Le usuali misure di asimmetria e di curtosi denotano una rapida convergenza alla gaussiana. Quindi possiamo considerare gaussiana la statistica

$$\frac{\nu - \dfrac{2(n-2)}{3}}{\sqrt{\dfrac{16n - 29}{90}}} \tag{5.21}$$

osservata in pratica. Se il suo valor p è inferiore a 0.01, si rifiuta l'ipotesi di casualità dei residui.

Test dei ranghi

Un altro *test* di erraticità che ricorre nello studio delle serie storiche è quello dei ranghi. Il confronto tra i residui osservati e l'ordinamento dei numeri naturali è esteso a tutte le coppie di errori consecutivi. In particolare, bisogna contare il

numero ν_1 di volte in cui si ha $a_j > a_i$ per $j > i$. Si considerano $n(n-1)/2$ coppie distinte. Il valore atteso in una serie del tutto erratica quale il *white noise* è

$$E(\nu) = \frac{n(n-1)}{4} \tag{5.22}$$

Valori di ν_1 molto superiori alla media fanno pensare ad un trend crescente, non del tutto rimosso dalla serie. Se invece ν_1 è molto minore della media il trend deve essere ritenuto decrescente. La statistica ν_1 è legata al famoso indice di correlazione per ranghi τ di Kendall

$$\tau = \frac{4\nu_1}{n(n-1)} - 1 \tag{5.23}$$

che varia tra -1 e 1 ed ha valore atteso nullo in una serie erratica. La sua deviazione standard è

$$\sigma_\tau = \sqrt{\frac{n(n+1)(2n+5)}{72}} \tag{5.24}$$

Il rapporto tra τ ed il suo scarto quadratico medio può essere usato per sottoporre a verifica l'ipotesi di casualità data la sua convergenza alla gaussiana standardizzata.

Test dei segni

Un altro utile criterio di accertamento della erraticità è il cosiddetto test dei segni. Ci si chiede se i segni dei residui del modello stimato si susseguano nel tempo in maniera casuale. Una coppia di segni successivi concordi: $(+,+)$ oppure $(-,-)$ costituisce una permanenza ed una variazione di segni successivi diversi $(+,-)$ oppure $(-,+)$ costituisce una discordanza. Ad esempio, la successione seguente:

\- \- \+ \- \+ \- \+ \+ \- \- \+ \+ \+ \- \- \+ \- \- \+ \+ \+ \+ \- \- \+ \+

Segni	+	−	*Totale*
+	7	6	13
−	7	5	12
Totale	14	11	25

Figura 5.2: *Test* dei segni

ha 25 coppie concatenate di segni a partire da $(-,-)$ e a finire con $(+,+)$. Le permanenze sono 12 e riportate sulla diagonale. Le discordanze risultano 13 e sono riportate fuori diagonale.

Se la serie fosse casuale la tabella dovrebbe mostrare indipendenza e quindi il confronto tra la tabella in caso di indipendenza e la tabelle delle frequenze degli

Segni	+	−	*Totale*
+	7	6	13
−	7	5	12
Totale	14	11	25

Segni	+	−	*Totale*
+	7.28	5.72	13
−	6.72	5.28	12
Totale	14	11	25

$$\chi_c^2 = \frac{(7-7.28)^2}{7.28} + \frac{(6-5.72)^2}{5.72} + \frac{(7-6.72)^2}{6.72} + \frac{(5-5.28)^2}{5.28} = 0.051$$

Figura 5.3: *Test* dei segni in caso di indipendenza

abinamenti dei segni effettivamente riscontrate dovrebbe fornire una diagnosi sulla erraticità dei residui. Nell'esempio considerato si è ottenuto un valor p di 0.8213. Se si afferma che la serie è casuale si è in errore 18 volte su cento. Se fosse vera l'ipotesi di casualità, un valore della statistica di 0.051 o più grande, lo si potrebbe osservare -per caso- il 18% delle volte cioè se si ritiene che circa una volta su cinque non è da considerarsi troppo raro, si può accettare l'ipotesi che i residui siano erratici.

Difference sign test

Si tratta di un test simile al precedente. Qui però si conta il numero di volte che due residui consecutivi hanno uno scarto positivo: $a_t - a_{t-1} > 0$. Indichiamo con ν_2 il numero di scarti positivi. In una sequenza del tutto erratica (del tipo *white noise*) i valori di ν_2 avrebbero valore atteso e varianza dati da

$$E(\nu_2) = \frac{n-1}{2} \qquad Var(\nu_2) = \frac{n+1}{12} \tag{5.25}$$

Per serie *white noise* molto estese, la distribuzione di ν tende alla gaussiana con

$$N\left(\frac{n-1}{2}, \sqrt{\frac{n+1}{12}}\right)$$

per cui un valore di ν molto superiore o molto inferiore alla media indica la presenza di un *trend* crescente o decrescente che deve essere scorporato dalla serie se si vuole applicare correttamente la metodologia Box-Jenkins.

5.2.4 Indici di gaussianità

Molte applicazioni statistiche, nei più svariati contesti, presuppongono che i residui dei modelli abbiano una distribuzione gaussiana o che si possa agevolmente ricondurre alla gaussiana: non a caso la sua denominazione più comune è "distribuzione

normale, lasciando intendere che sia usuale osservarla. Non è così. Tale distribuzione è evocata allorché si debbano trattare questioni di inferenza concernenti la significatività dei parametri stimati, la loro stima intervallare, gli intervalli di previsione sui valori futuri del modello, i test di adattamento quali quelli considerati nel paragrafo precedente. Inoltre, la gaussianità è una ipotesi quando si intende pervenire a stimatori corretti ed a varianza minima.

Ci sono due buone ragioni per spiegare la diffusione e la prevalenza della distribuzione gaussiana per approssimare le distribuzioni riscontrare empiricamente. Innanzitutto la grande generalità del teorema limite centrale. La sua affermazione principale è che la distribuzione della somma di un numero elevato di variabili casuali indipendenti e identicamente distribuite e con variabilità contenuta tende a distribuirsi secondo la gaussiana, indipendentemente dalla distribuzione delle singole variabili. La seconda ragione è la coincidenza che, in caso di gaussianità, si determina tra il metodo dei minimi quadrati ed il metodo della massima verosimiglianza. Tale identità è particolarmente rilevante dato che rende le stime ai minimi quadrati *best linear unbiased estimator* per i parametri incogniti laddove l'efficienza dei minimi quadrati è molto meno elevata qualora le osservazioni abbiano una distribuzione asimmetrica, bimodale o a code pesanti o una combinazione di queste caratteristiche.

Alcune considerazioni si impongono. Il teorema del limite centrale può essere inapplicabile dato che le distribuzioni reali sono raramente indipendenti o identicamente distribuite o con varianza ridotta. In pratica, non è affatto "normale ottenere dati che seguono la distribuione normale. Le banche, ad esempio, analizzano milioni di conti correnti o di deposito ed ognuno potrebbe essere descritto da una distinta variabile casuale, ma non è pensabile che il comportamento di ciascuno si realizzi ignorando quello degli altri. Rispetto ad un livello medio non è giustificabile l'ipotesi che gli scostamenti in più o in meno abbiano la stessa probabilità di verificarsi: più facilmente si verificheranno rari episodi di depositi elevati e poi piccoli depositi determinando le code pesanti tipiche dei dati finanziari. In questi casi mancano le condizioni basilari del teorema del limice centrale. Infatti, i modelli stocastici della volatilità convergono molto lentamente alla gaussianità.

Nelle applicazioni del mondo reale (non didattico o accademico) si è quasi sempre incerti su quale sia la vera distribuzione dei residui o perché si hanno pochissimi dati su cui ragionare oppure perché non si ha alcuna idea di quale possa essere tale distribuzione oppure perché ne sono noto alcune caratteristiche tendenziali senza che però si riesca a formulare una espressione precisa. In questi casi è facile cedere alla tentazione di proporre comunque la distribuzione gaussiana come modello di riferi-

mento e quindi mutuare, senza giustificato motivo, risultati che sarebbero validi forse asintoticamente, come caratteristiche campionarie per le dimensioni con cui si sta trattando. Giova ricordare che non sempre è utile cercare trasformazioni o attuare cancellazioni nei dati al fine di ottenere una distribuzione accettabilmente gaussiana perché nella maggior parte dei casi, i dati non obbedicono a questo modello. E' vero che presupponendo la gaussiana diventano subito disponibili i valor p di tutti i test per valutare la qualità del modello. Ma se non ci sono ragioni specifiche di adottarlo è meglio ricorrere a distribuzioni più flessibili. Il fatto che il modello gaussiano sia troppo buono per non essere quello giusto non è un ragionemento scientifico.

L'importanza del modello gaussiano è tale che ben pochi problemi statistici hanno ricevuto più attenzione di quella dedicata all'accertamento della gaussianità. Infatti, uno dei test che lega il momento terzo ed il momento quarto del campione dei dati ha radici che affondano negli anni venti del secolo scorso. Si tratta del test di Jarque-Bera (1980) e Bowman-Shenton (1975)

$$JB = n\left[\frac{\left(\sqrt{\gamma_1}\right)^2}{6} + \frac{(\gamma_2 - 3)^2}{24}\right] \sim \chi^2(2) \tag{5.26}$$

dove n è la lunghezza della serie storica, $\sqrt{\gamma_1} = m_3/(m_2)^{1.5}$, $\gamma_2 = m_4/(m_2)^2$, ed $m_i, i = 2, \ldots, 4$ sono i momenti centrali cioè

$$m_i = \frac{\sum_{t=1}^{n}(y_t - m_1)^i}{n}; \quad m_1 = \frac{\sum_{t=1}^{n} y_t}{n} \tag{5.27}$$

In breve, il test JB, verifica la gaussianità accertando che il momento terzo sia nullo $m_3 = 0$ e che per il momento quarto si abbia $m_4 = 3m_2^2$ che è quanto succede con la gaussiana. E' ovvio che scostamenti dalla gaussiantià che non si riflettono nel terzo e nel quarto momento non possono essere rilevati dalla JB. Il software per l'approccio Box-Jenkins fornisce il p-value di questo test che tuttavia rimane poco attendibile per serie storiche inferiori alle 100 osservazioni a causa della lenta convergenza alla gaussianità (e quindi al chi quadrato) delle quantità coinvolte nella statistica JB. Inoltre, il test JB non è applicabile se nei residui permane della autocorrelazione.

Un altro *test* basato sulla curtosi è quello di Geary ottenuto con il rapporto tra la deviazione media (media degli scarti assoluti dalla media aritmetica) e lo scarto quadratico medio:

$$g = \frac{\sum_{t=1}^{n}|y_t - m_1|}{\sqrt{n\sum_{t=1}^{n}(y_t - m_1)^2}} \tag{5.28}$$

In una distribuzione gaussiana questo rapporto è pari a $\sqrt{2/\pi}$ ed ogni allontanamento da questa soglia potrà essere considerato un indizio di disnormalità. Il test

basato su g è meno efficace dello JB nel senso che rigetta l'ipotesi nulla meno del dovuto quando essa è sbagliata (g ha minore potenza rispetto a JB); tuttavia converge più rapidamente alla Gaussiana e quindi i suoi valor p sono più vicini alla realtà per serie storiche più brevi. Urzua (2007) ha coinvolto la statistica di Geary sulla curtosi con la componente di asimmetria del test di JB ottenendo la statistica

$$Q = \left[\frac{m_3}{dm_2^{1.5}}\right]^2 + \frac{(w-3)^2}{e} \tag{5.29}$$

dove

$$d = \frac{6\,(n-2)}{(n+1)\,(n+3)}; \quad e = \frac{3.54}{n+2}; \quad w = \frac{-6\ln(g)}{\ln(\pi/2)} \tag{5.30}$$

La Q ha distribuzione asintotica chi-quadro con due gradi di libertà.

Esistono molti altri test di gaussianità: test di Kolmogorv-Smirnov, test di Shapiro-Wilk, test di D'Agostino-Pearson, etc. Si veda [**?**]. Alcuni considerano i valori osservati dei residui un quasiasi campione di osservazioni indipendenti e che quindi potrebbe essere permutato senza alcuna conseguenza per l'efficacia del test; altri invece prendono in considerazione esplicitamente la natura di serie storica dei residui e, pertanto, dovrebbero risultare più attendibili nella verifica della gaussianità.

L'utilità dei grafici è ben nota. Di seguito presentiamo l'istogramma dei residui ed il grafico QQ (quantile-quantile) basato sulle statistiche d'ordine della gaussiana.

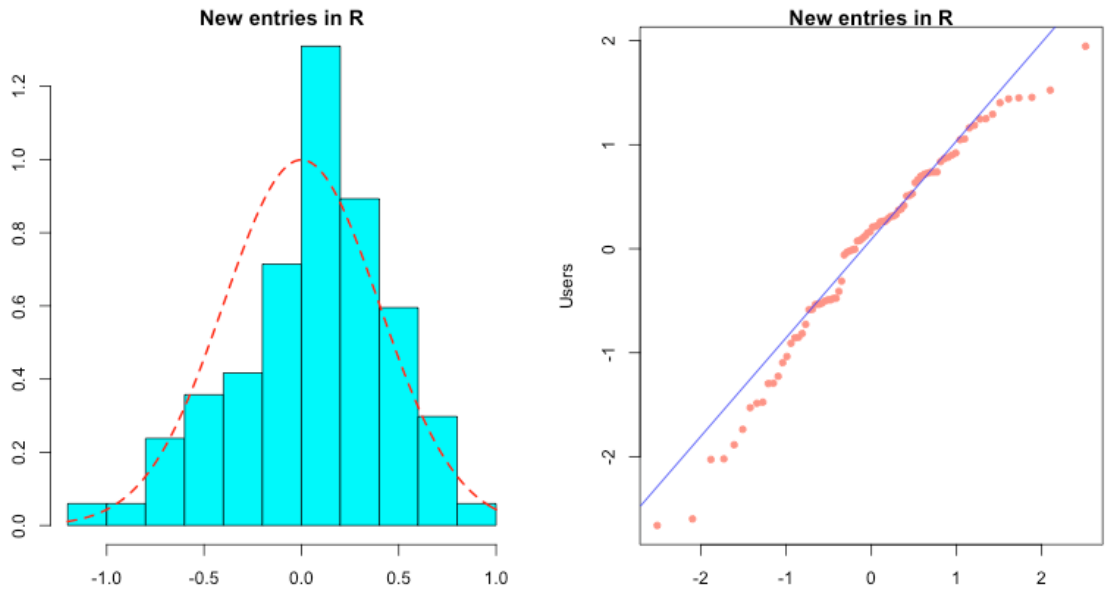


Figura 5.4: Grafici per valutare la gaussianità dei residui

5.3 Identificazione

I manuali sulle serie storiche editi fino alla metà degli anni '90 sono ricchi di suggerimenti su come procedere alla scelta degli ordini dei polinomi caratteristici p, P, q, Q che dia il miglior adattamento ai dati osservati con parametri significativi e residui *white noise*. I suggerimenti sono intrisi di un approccio artigianale, passo-passo con approfondimenti, modifiche, analisi grafiche, ripensamenti e faticosi calcoli da fare e rifare con scarsa interazione. Insomma, come un pittore davanti ad un quadro importante.

La operatività della tecnica Box-Jenkins prevedeva tre stadi successivi per arrivare al modello definitivo più consono alla serie storica osservata.

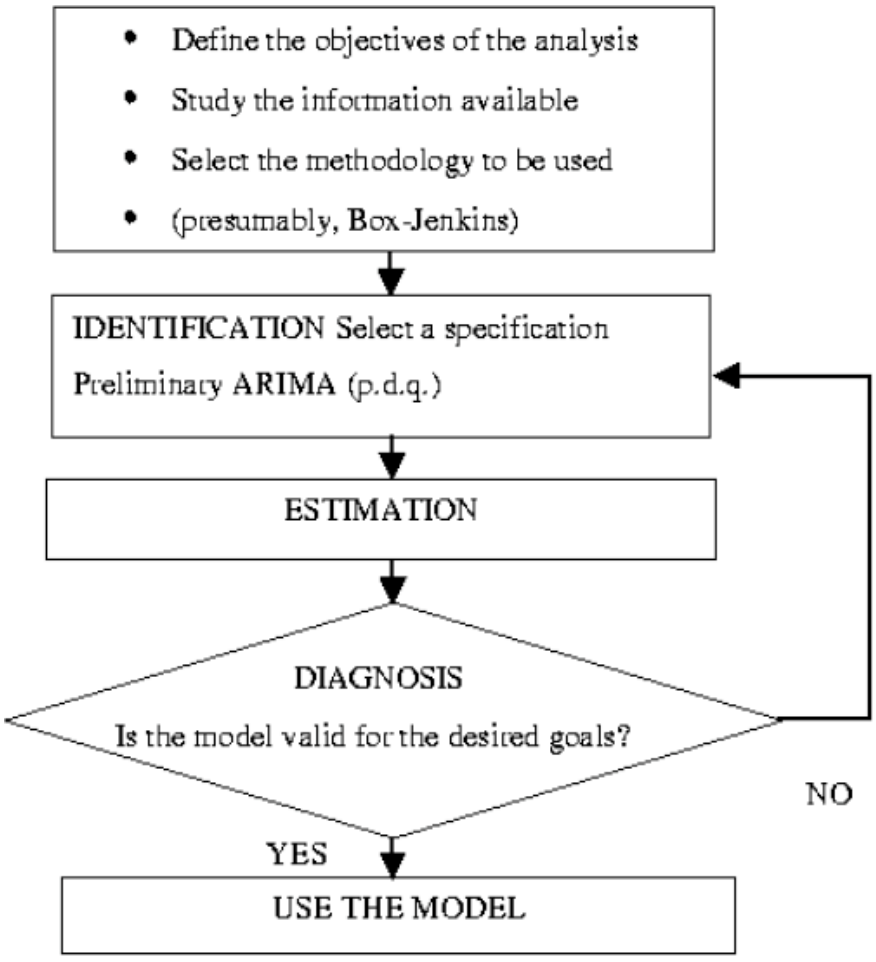


Figura 5.5: Schema tradizionale della procedura BJ

L'analisi grafica include la ricerca del *trend* nonché la verifica della stagionalità e quindi dell'ordine delle differenze regolari e stagionali. In questa fase si colloca lo studio dei correlogrammi, semplici e parziali, per stabilire l'ordine del modello

regolare ed eventualmente di quello stagionale. La stima prevede il calcolo dei coefficienti e la valutazione dei t-Student. Nella valutazione delle diagnostiche è previsto l'insieme dei controlli da effettuare sui residui che sono setacciati per valutare la rimanenza di sistematicità, adattamento del modello, erraticità dei residui. L'analisi porta al giudizio di qualità sul modello proposto: se il giudizio è positivo si procede alle previsioni, se è negativo si esamina un altro modello. Infatti, la ripetizione del ciclo dei tre stadi: scelta del modello, stima e diagnostica è caratteristica di questo approccio e, alla fine del ciclo ci si può trovare con più di una proposta valida.

La visione artistica per molti versi intrinseca alla metodologia BJ non poteva reggere alla "rivoluzione industriale" avvenuta nell'analisi dei dati che ha portato all'attenzione diffusa una mole enorme di dati ed alla esigenza di trattare tantissime serie storiche relative a fenomeni distinti e distanti, in tempi ristretti e con professionalità solo parzialmente disponibili. La metodologia BJ ha infatti avuto solo una tiepida accoglienza nel mondo delle aziende e questo è da attribuire alle difficoltà di identificare e poi usare un buon modello con questa tecnica.

E' chiaro che un giudizio positivo sul modello sotto esame non può escludere che un altro modello non sia altrettanto valido se non migliore. Tuttavia, la procedura BJ effettuata con questo *iter* è troppo lenta e dispersiva nonché soggetta a molte arbitrarietà e richiede importanti competenze specifiche nelle serie storiche.

L'evoluzione dei mezzi di calcolo consente elaborazioni impensabili negli anni '70 allorché le tecniche dei processi stocastici lineari cominciarono a consolidarsi. Nel frattempo sono migliorate le procedure di acquisizione dei dati che si sono in buona parte automatizzate e quindi depurate dagli errori umani più perniciosi. D'altra parte si sono estesi i campi tradizionali di applicazione dell'analisi delle serie storiche che è andata ad innervare le tecniche quantitative di moltre altre discipline quale la medicina, la robotica, i trasporti, i servizi di sicurezza. Questa premessa serve a legittimare, almeno in parte, uno schema operativo in cui il ruolo dell'agente umano è ridotto in termini di impegno "esecutivo", ma è aumentato per gli aspetti di indirizzo e controllo che potrebbe portare alla scissione dei due momenti di ricerca con prevalenza netta del secondo sul primo.

La questione può essere così sintetizzata: sono disponibili i valori di una serie storica formatisi nella rilevazione su scala metrica o intervallare di un certo fenomeno che si evolve nel tempo. Secondo la nostra impostazione, le analisi preliminari si effettuano prima di procedere alla applicazione della metodologia Box-Jenkins: quando le serie storica arriva ai modelli ARIMA, si presume che siano già state effettuate le operazioni preliminari *standard* e cioè le correzioni di calendario per rendere uniformemente comparabili le unità di rilevazione, le periodicità stagionali

siano state confermate o sconfermate, i *remota* siano stati sostituiti se manifestamente sbagliati, i valori mancanti siano stati imputati con pseudo osservazioni plausili. Si presumono pure effettuare le analisi per segmenti per asseverare l'assenza di forte eteroschedasticità nonché l'assenza di eventuali cesure storiche (fratture del ciclo-trend).

Lo scenario che si è formato nel corso degli ultimi anni, anche grazie alla disponibilità di software gratuito, sofisticato al punto giusto e, soprattutto, attendibile è quello in cui si sottopongono a stima e verifica automatica centinaia se non migliaia o decine di migliaia di modelli alternativi per ciascuna serie storica. Tra questi si chiede al software di evidenziarne solo alcuni, avendone il tempo, per una analisi di dettaglio. Se serve, si può procedere a selezionare il migliore, impiegandolo senza indugi nelle previsioni.

Sebbene questa impostazione urti la sensibilità dei professionisti del campo, la selezione automatica del modello si è arricchita di paradigmi mutuati dall'intelligenza artificiale: logica *fuzzy*, algoritmi genetici, etc. riducendo o eliminando l'intervento dell'agente umano. A questo approccio si è arrivati grazie all'aumento delle attività nrl campo delle serie storiche che non solo ha velocizzato le tecniche esistenti ed agevolato il lavoro di chi già operava in questo campo, ma ha ampliato l'opportunità di usare la BJ anche a molte altre discipline e nuovi studiosi.

T.J. Watson, presidente della IBM, nel lontano 1943 affermò : *"I think there is a world market for about five computers"*. Bill Gates, CEO della Microsoft nel 1981 predisse che un disco rigido da mezzo mega avrebbe dovuto essere sufficiente per chiunque. Qual'è l'errore? Secondo [**?**] è quello di concentrarsi sulle attività correnti e non di altre attività, del tutto nuove, che possono avvantaggiarsi, dalla nuova tecnologia.

5.3.1 Identificazione automatica

Il livello del supporto che possono dare i mezzi di calcolo, almeno quelli cui può accedere un singolo studioso o professionista ai tempi in cui queste note sono redatte, richiede la definizione di un ambito in cui effettaure la ricerca al fine di evitare la inutile disamina di molti modelli inadeguati. Ad esempio, la trasformazione Box-Cox e la differenziazione -regolare o stagionale- devono essere fissate a priori oppure sono da considerarsi "parametri" incogniti da determinare insieme agli altri? Nei programmi che accompagnano il presente testo, la Box-Cox è stata ridotta alla scelta tra la scala logaritmica e la scala naturale ed è considerata un termine di ricerca. Così è pure per le differenze.

La strategia proposta è di stabilire una classe di modelli per le serie regolari ed una classe di modelli per le serie stagionali nell'arco dei termini seguenti

$$\begin{aligned} &\text{serie regolare :} \quad ARIMA\,(0:12, 0:2, 0:6) \\ &\text{serie stagionale :} \quad ARIMA\,(0:3, 0:2, 0:3) \times ARIMA\,(0:2, 0:1, 0:2)_s \end{aligned} \tag{5.31}$$

che comportano la disamina di 546 modelli di serie regolari (di cui la metà in scala logaritmica) e 862 per le serie stagionali. Non è inopportuno ricordare, come termine di paragone, che [**?**] proponeva la scelta di tre modelli su 12 potenziali. In questo caso, poiché la varianza in scala logaritmica è inferiore a quella in scala naturale conviene inserire nelle misure di adeguatezza una qualche quantità che ne corregga la scala. Nei criteri indicati nella relazione (5.15) abbiamo rapportato la stima della varianza degli errori alla varianza della serie originale. Ad esempio:

$$\text{Akaike information criterion (AIC): } C = -\log\left[\frac{\hat{\sigma}_a^2}{\sigma_w^2}\right] + \frac{2m}{n} \tag{5.32}$$

Sembra tutto a posto, ma non è così. Potrebbe insorgere la necessità di ampliare il campo di ricerca o di dover ripetere l'intera procedura tenendo conto del trattamento dei valori remoti e dei valori mancanti o di una trasformazione migliore di quella logaritmica per ridurre l'eteroschedasticità. Qual'è la direzione da seguire? E più prudente favorire i modelli maggiormente parsimoniosi collocandoli nelle prime posizioni tra le formulazioni da prescegliere oppure è più saggio assecondare la scelta di modelli con le migliori diagnostiche?

C'è poi un altro problema: la distensione delle serie storiche può arrivare agevolmente alle centinaia di migliaia di valori ed i tempi della ricerca automatica potrebbero risentirne al punto tale da restringere i campo di ricerca ai pochi modelli anticamente raccomandati da Box e Jenkins. L'architettura attuale dei microprocessori non consente troppe variazioni sulle serie storiche lunghissime o su modelli di ordini molto elevati, ma è solo questione di tempo, non di metodo.

Studiamo come esempio la serie dei *motor vehicles engines*. La ricerca automatica implementata nella procedura *auto.arima* del pacchetto *forecast* individua un modello ARIMA$(1,0,1) \times (1,0,1)_{12}$

$$W_t = \phi_1 W_t + \Phi_1 W_{t-12} + a_t - \theta_1 a_{t-1} - \Theta_1 a_{t-12} \tag{5.33}$$

la cui stima produce i risultati che seguono:
La nostra tecnica, che considera tutti i modelli indicati nella relazione (5.31), trova invece delle soluzioni aggiuntive elencate nella figura 5.7 dove, in particolare la

```
Coefficients:
         ar1      ma1    sar1     sma1
      0.9628  -0.3931  0.9739  -0.8185
s.e.  0.0200   0.0857  0.0240   0.0817

sigma^2 estimated as 65106:  log likelihood = -1315.09
AIC = 2640.17   AICc = 2640.5   BIC = 2656.35
```

Figura 5.6: Identificazione automatica della serie *motor vehicles engines*

```
          [,1] [,2] [,3] [,4] [,5] [,6] [,7]
[1,] 2457.222    3    1    2    0    1    1
[2,] 2460.031    4    1    2    1    1    1
[3,] 2460.715    3    1    2    1    1    1
[4,] 2461.438    4    1    1    1    1    1
[5,] 2461.960    4    1    1    0    1    1
[6,] 2462.803    3    1    2    2    1    1
[7,] 2462.883    2    1    1    1    1    1
```

Figura 5.7: Stima del modello identificato

```
Coefficients:
         ar1      ar2      ar3      ma1     ma2     sma1
      1.4181  -0.2944  -0.2784  -1.9144  0.9998  -0.8452
s.e.  0.0747   0.1301   0.0752   0.0760  0.0793   0.0652

sigma^2 estimated as 61446:  log likelihood = -1222.3
AIC = 2456.6   AICc = 2457.27   BIC = 2478.75
```

Figura 5.8: Identificazione automatica della serie *motor vehicles engines*

prima sembra rispondere a tutte le nostre aspettative. E' su questa configurazione: ARIMA$(3,1,2) \times (0,1,1)_{12}$ che concentreremo la validazione e le previsioni.

I parametri sono significativi, tranne che per il termine AR2 della parte regolare che però ha un t-Student pari a $t = 1.88$ che non è proprio da buttare. I parametri della componente MA della parte regolare sono, in valore assoluto, maggiori di uno ($\hat{\theta}_1 = -1.94$) o quasi uguali ad uno ($\hat{\theta}_2 = 0.9998$) che pongono in forte dubbio l'invertibilità e quindi l'identificabilià del processo retrostante la serie storica.

Per completezza conviene adattare alla serie anche un modello che abbia solo la parte autoregressiva, ma con un ordine molto elevato rispetto a quello che è usuale nei modelli ARMA misti. L'idea è che, siccome ogni ARMA può essere ricondotto ad un AR con ordine infinito, possiamo tentare di approssimare l'ARMA insoddisfacente con un AR finito, ma a con un ordine elevato
Emerge la proposta di un ARIMA$(25,0,0)$ in cui però solo i ritardi $1, 7, 13, 25$ sono significativi. Il modello risulta stazionario. Saranno i successivi accertamenti a guidare la scelta definitiva. L'analisi dei residui del primo modello ARIMA$(3,1,2) \times$

```
> FitFDA
AR(25). MLE.  Mean estimated using the sample mean
length of series = 188 ,  number of parameters = 5
loglikelihood = -1055.973 ,  AIC = 2121.9 ,  BIC =  2138.1 , UBIC =  2159.9
> coef(FitFDA)
                 MLE          sd   Z-ratio
zeta(1)     0.8405338  0.02961990 28.377335
zeta(7)     0.2110509  0.08947002  2.358900
zeta(13)   -0.4047635  0.08228466 -4.919064
zeta(25)   -0.2587939  0.07044787 -3.673552
mu       1343.6880702 88.41621078 15.197304
```

Figura 5.9: Identificazione automatica e stima di un AR ad alto ritardo

$(0,1,1)_{12}$ produce il seguente grafico

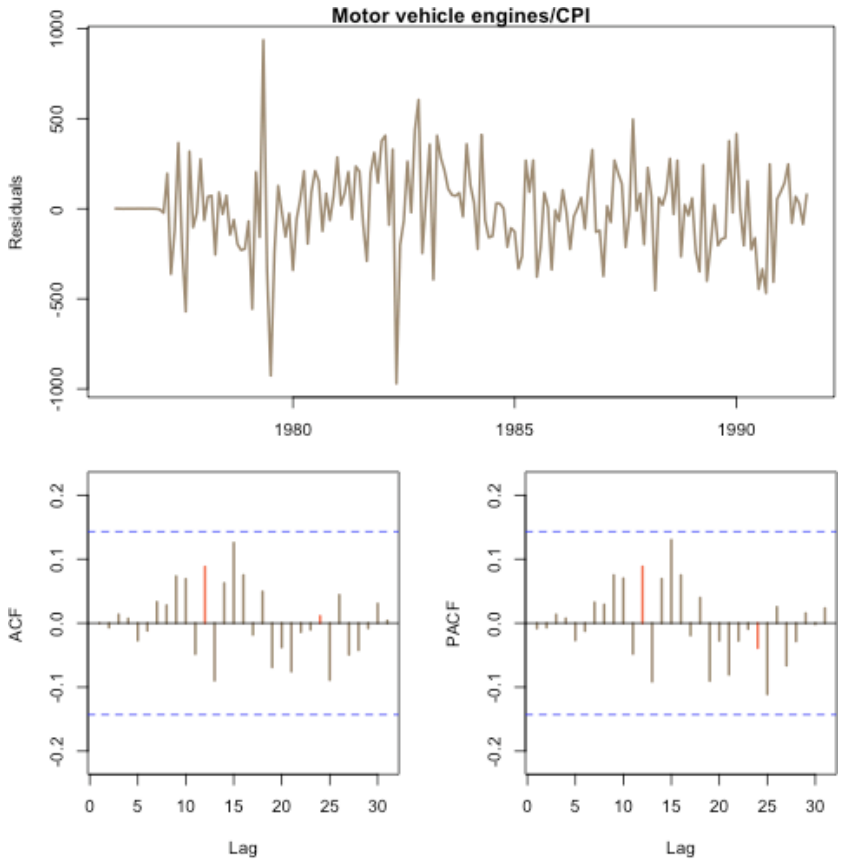


Figura 5.10: Analisi della correlazione nei resiui per ARIMA$(3,1,2) \times (0,1,1)_{12}$

in cui non si notano autocorrelazioni semplici o parziali significative o presumibilmente tali. Anche i test di casualità dei residui confermano che in essi non è rimasto alcunché di utilizzabile ai fini del riconoscimento del modello.

La stima del secondo modello determina i residui riportati nella figura 5.12. Le autocorrelazioni sembrano tutte non significative collocandosi al di sopra del livello

	Statistic	p-value
Ljiung-Box statistic	18.00505	0.84218
Modified LB test	20.57200	0.71621
Monti test:	20.29269	0.73131
Turning points	0.11619	0.90750
Runs test	0.07667	0.93889
Rank Test	-0.00230	0.99817
Sign test	93.00000	0.87596

Figura 5.11: Analisi della casualità dei residui per ARIMA$(3,1,2) \times (0,1,1)_{12}$

di soglia de 5%. Forse questo ARMA è una alternativa praticabile

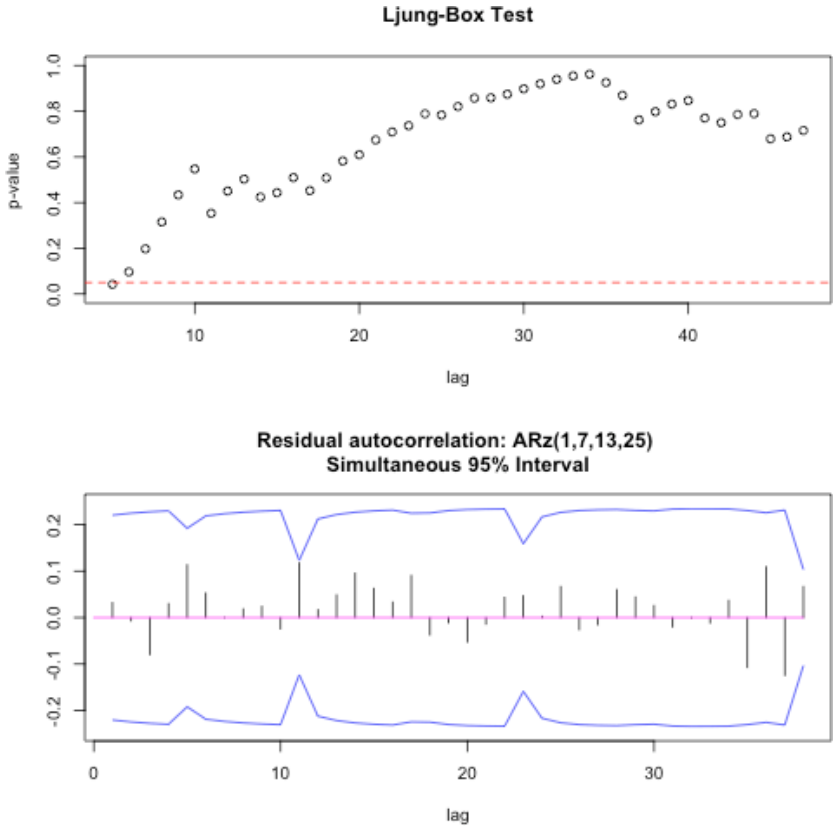


Figura 5.12: Analisi della correlazione nei residui per ARIMA$(25,0,0)$

La figura 5.13 induce a giudicare sospetta la prima autocorrelazione parziale anche se il valore è veramente esiguo. Tuttavia questa presenza, per quanto evanescente, affiora nei test di casualità del modello a grande ritardo perché almeno tre dei test considerati hanno un valor p più basso di quanto sarebbe gradito per un modello ben fatto. Attenzione! Si tratta del Monti test basato proprio sulle autocorrelazioni parziali e del test dei punti di svolta legato all'aucorrelazione di ritardo uno e quindi alla prima autocorrelazione parziale. Non possiamo quindi trascurare tutti questi segnali: sono singolarmente flebili, ma insieme ci avvertono che un modello formato

solo dalla parte AR non è proponibile dato che lascia nei residui una parte che invece avrebbe dovuto assorbire. In effetti, anche dopo aver incluso altri parametri autoregressivi, il difetto è rimasto e questo significa che si tratta di un segno che è tipico di questa applicazione se non proprio del metodo contenuto nel pacchetto *FitAR*.

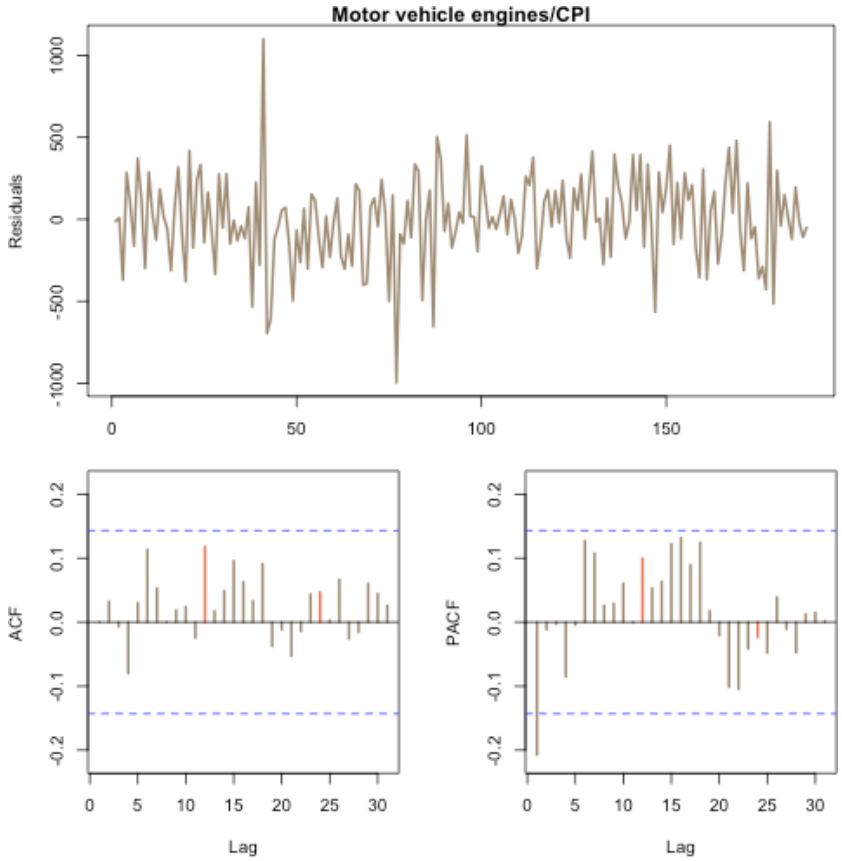


Figura 5.13: Ulteriore analisi della correlazione nei residui per ARIMA$(25, 0, 0)$

	Statistic	p-value
Ljiung-Box statistic	25.66085	0.42587
Modified LB test	28.18659	0.29934
Monti test:	37.56451	0.05099
Turning points	2.38187	0.01722
Runs test	2.63262	0.00847
Rank Test	-0.00230	0.99816
Sign test	90.00000	0.13623

Figura 5.14: Analisi della casualità nei residui per ARIMA$(25, 0, 0)$

5.4 Possibilità e limiti delle previsioni

Uno degli obiettivi della metodologi BJ è di ottenere previsioni attendibili per i valori futuri della serie storica di modo che chi prende decisioni può ancorarle, se lo ritiete utile ed opportuno, ad una base informativa più ricca e meglio organizzata.

Gli aruspici predicevano il destino studiando attentamente il fegato e l'intestino (soprattutto fegato) degli animali sacrificati (in genere pecore): se osservavano segni particolari come cicatrici o altre anomalie, confrontavano il fegato con un modello bronzeo (famoso è il Fegato di Piacenza, modello in bronzo riportante le ripartizioni e i nomi degli dèi) per capire a quale settore del cielo corrispondeva e, quindi, quale divinità avesse mandato quel segno (se era di buon auspicio o meno), per poi cercare di capirne il significato. Sebbene i moderni esperti di statistica, econometria, mercati finanziari, etc, siano in alcuni casi considerati degli aruspici senza animali da scannare, le previsioni scientifiche riescono almeno a fare previsioni valide se si ragiona su scale temporali più ampie e fenomeni più diffusi.

L'opinione incerta che domina sulle previsioni scientifiche relative ai fatti che ci riguardano più da vicino: prezzi, occupazione, rendimenti ha origine nelle attese eccessive che si ripone in strumenti che, per natura, non possono essere precisi. Secondo Frain (1992), le previsioni scientifihce perseguono due finalità

1. Stabilire un valore probabile o atteso di una quache grandezza di interessa, ad esempio, il PIL in un periodo futuro.
2. Ridurre l'ampiezza dell'arco dei valori che saranno raggiunti da un certo evento in un futuro più o meno lontano.

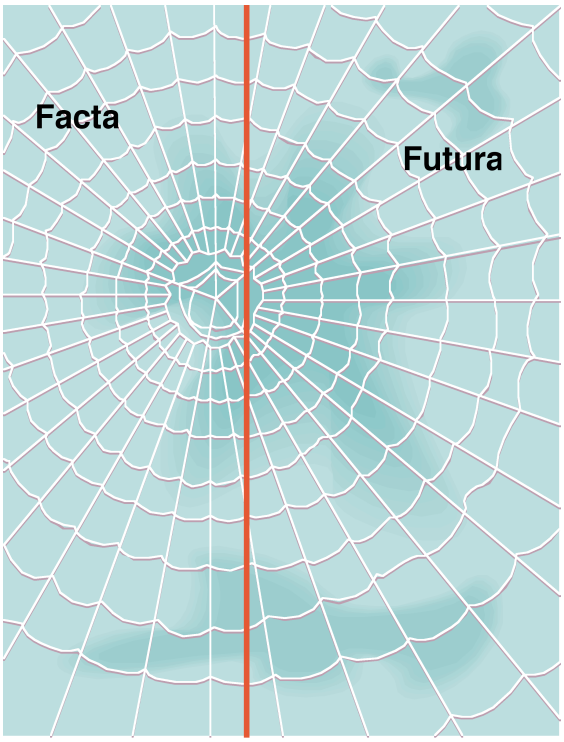
"Valore atteso","Riduzione dell'incertezza" sono locuzioni che debbono fare intendere che, decidere sulla base di previsioni, per quanto scientifiche, comporta un certo rischio cioè non è possibile conoscere con certezza le premesse su cui la decisione di basa e nemmeno coscere con certezza le conseguenze della decisione.

Il Rischio è, in fondo, ignoranza del futuro: l'onniscienza, la facoltà di essere a conoscenza di tutto, esclude ogni forma di rischio razionale. Le previsioni scientifiche, sfruttando bene le osservazioni già fatte, aumentano le nostre conoscenze sul futuro e possono ridurre il rischio; ridurre, non eliminare perché il rischio rientra nella vita di tutti coloro che onniscienti non sono. Le assicurazioni non eliminano il rischio: trasferiscono, ad un certo costo, le conseguenze del rischio su qualcun altro.

Distinguiamo gli eventi in due tipi:

- **Facta** - avvenimenti che si collocano nel passato.
- **Futura** - avvenimenti che si collocano nell'avvenire

Il loro punto di incontro è quell'istante brevissimo in cui la delusione cede il passo alla speranza. In quanto soggetti agenti possiamo intervenire solo sui **Futura**: se qualcosa non è ancora accaduto è almeno concepibile l'idea di agire per modificarne il corso. D'altra parte, è possibile conoscere e descrivere solo i **Facta**. Ciò che sarà può solo essere prefigurato: profezie, congetture, prospettive, proiezioni, etc.



Immaginiamo il sistema degli eventi, passati e futuri, come una ragnatela. Se la si tocca in un punto ci saranno reazioni moltiplicative in tutti i fili. Il sistema degli eventi futuri è visto come un complemento di quello del passato. Attraverso la conoscenza delle reazioni in questo sia possibili stabilire le reazioni sui fili del futuro. Con una sola grande precisazione: la conoscenza, anche accurata del passato, non potrà mai configurare fatti ed eventi del tutto nuovi che avverranno nel futuro e non sono mai avvenuti prima, neanche in forma embrionale. Le crisi dei mercati finanziari del 2008 o del 1987 non sono state previste. Certo, alcuni sostengono di averle annunciate e forse è vero: per fortuna, perché avevano accesso ad informazioni giuste, ma riservate, perché a predirne tante qualcuna si azzecca (oroscopi, tarocchi, lettura della mano); si può però escludere che il verificarsi di fenomeni così devastanti possano essere predetti da modelli di previsioni scientifiche.

La previsione è la descrizione di eventi futuri che si basa sulla conoscenza di eventi passati e su di un insieme di ipotesi. In generale, una tecnica di previsione

fornirà il valore atteso, in base al quadro di conoscenze disponibili, con l'aggiunta di indicazioni sull'incertezza della previsione che usualmente si condensa in uno o più intervalli all'interno dei quali il valore incognito futuro tenderà a collocarsi (intervalli di previsione).

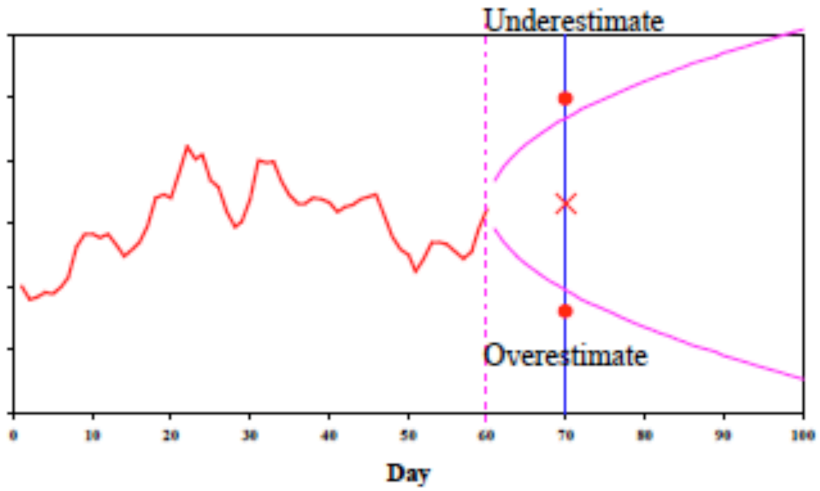


Figura 5.15: Schema generale delle previsioni

Questo è frequente anche nel linguaggio comune: sarò in sede alle cinque, minuto più, minuto meno tuttavia non tutte le tecniche di previsione sono in grado di produrre intervalli utili e coerenti con le premesse alla base della tecnica stessa; ad esempio, per il livellamento esponenziale si ottengono gli intervalli di previsione sfruttando le analogie e le similitudini con altre tecniche. (Si veda [**?**]).

Secondo N. Meade la previsione non dobrebbe riflettere l'intuito influenzabile delle persone: le tecniche automatiche di previsione funzionano meglio, a parità di informazioni. In effetti, noi tendiamo ad essere abbagliati dagli eventi estremi, particolarmente quelli recenti: gli ottimisti, ad esempio, confondono spesso gli obiettivi con le previsioni cioè è il desiderio di un risultato che porta a prevedere quel risultato (che talvolta, specie nei mercati finanziari, provoca il risultato); i pessimisti tendono a volare raso terra dopo anche un piccolo inconveniente. Il famoso Bill Gates, agli inizi della sua attività nel 1981 affermò che un disco rigido di $640k$ sarebbe stato sufficiente per chiunque; oggi un computer di basso livello ne ha per un miliardo di volte in più.

D'altra parte tendiamo a dimenticare gli errori di previsione commessi in passato dando invece un peso straordinario ai purtroppo rari successi. Siamo poi bravissimi, ad evento accaduto, ad affermare: l'avevo detto (tecnica del "senno di poi"). Le previsioni sono un'attività troppo seria per non affidarla alle macchine Al 17 ottobre 1929, l'indice azionario aveva raggiunto un *plateau* elevato e permanente prima di



Figura 5.16: Esempio di ottimismo seriale (Irvine Fisher, Yale)

un rovescio durato final 14 novembre 1929 che lasciò sul terreno il 36% del valore. Una caduta che era comunque destinata ad esaurirsi in poco tempo, pochi giorni al massimo e la situazione si sarebbe stabilizzata in alcuni mesi (nelle previsioni o nei desideri?). Dopo tre mesi vi fu il famoso crollo di *Wall street* e l'indice non si riprese prima del tardo 1936. L'incapacità di prevedere punti di rottura catastrofici è intrinseca a molti dei modelli usati per le previsioni scientifiche. Infatti, una delle ipotesi su cui si fondano è che, nel periodo in esame non ci saranno disastri. L'economista Fisher forse non aveva le categorie per inquadrare così rilevanti mai avvenuti prima, ma rimane il sospetto che abbia agito da ottimista seriale, poco incline ad imparare dai propri errori.

Gli errori di previsione sono inevitabili, ma con il procedere della evoluzione del fenomeno gli errori dovrebbero attenuarsi se la tecnica di previsione è corretta e se il fenomeno non devia troppo dalla struttura già seguita. Quello che differenzia le previsioni scientifiche dalla sfera di cristallo della Megera è che i primi si sforzano di incorporare gli errori nel meccanismo di previsione in modo che, almeno tendenzialmente, i prossimi errori saranno minori. Non esiste la fortuna, esiste il momento in cui il talento incontra l'occasione.

La previsione altro non è che l'estensione ragionata ai **Futura** delle regolarità accertate per i **Facta** in base uno schema condiviso di rilevazione. Su quali basi logiche si fonda tale estensione? Lo schema è riconducibile alla induzione per enumerazione: Dato che per certo numero, dicamo n di volte si è trovato che gli oggetti α sono dei β e che in nessuno di questi casi si è trovato che un β non fosse un α, le due affermazioni:

- Il prossimo α sarà anche un β.
- Tutti gli α sono anche dei β.

diventano sempre più verosimili man mano che n si avvicina all'infinito. Comunque, il principio dell'induzione non può essere generalizato perché si rivela spesso falso: una sequenza di successi in un campo autorizza a sperare che il prossimo evento sarà un successo, ma non può darne la garanzia. Possiamo però seguire il principio di induzione in questo senso. L'aver constatato che certi eventi si sono manifestati finora in una data maniera non autorizza a dedurne che essi continueranno a presentarsi nel medesimo modo in futuro. Nulla però vieta di pensare che, fra tutte le possibili manifestazioni, la più plausibile sia proprio la ripetizione nei modi e nelle forme consuete. Anzi, tale metodo della **Previsione Annunciata**, è il criterio tipico adottato nelle scienze empiriche: in presenza di certe condizioni si prevede il verificarsi di un evento; se questo succede, si ha una conferma, ma sono richiesti ulteriori accertamenti. Più accertamenti arrivano, più sicuri siamo della previsione.

Il rischio connesso con questa impostazione è illustrato magnificamente in [**?**] che spiega la previsione per una serie storia univariata come se si stesse guidando con gli occhi fissi sullo specchietto retrovisore



Figura 5.17: Guida spericolata

In prima istanza, tutto continuerà a svolgersi secondo il tratto di strada già percorso. Ma fino a quando? se la strada è in salita ci aspettiamo una discesa, ma quanto tempo dobbiamo aspettare?

5.4.1 Le previsioni nelle scienze sociali

Esistono eventi caratterizzati da molta stabilità nel loro manifestarsi, ma ne esistono altri assai più mutevoli.

- I primi sono frequenti nelle scienze fisiche.
- I secondi sono frequenti nelle scienze sociali.

La loro prevedibilità è molto ineguale: E' retorico domandarsi se un oggetto, lasciato libero, cadrà oppure rimarrà sospeso nell'aria in presenza di gravità. L'idea invece che un aumento di prezzo induca una contrazione della domanda è solo plausibile. Le tendenze riscontrate nelle scienze sociali sono più imprecise e vaghe che nelle scienze fisiche. In molti contesti sono talmente incerte da far dubitare persino della loro utilità.

Le previsioni in campo economico sono così distinguibili:

(a) A brevissimo termine (fino a ad un mese). In questo ambito rientrano le previsioni JIT (*just-in-time*) ed il cosiddetto *nowcasting* in cui l'orizzonte temporale può essere la mezzora o i cinque minuti come nel caso della richiesta di energia elettrica o di inquinamento ambientale.

(b) A breve termine (fino ad un anno). Hanno un ruolo di primo piano i fattori accidentali o erratici: scioperi, leggi speciali, catastrofi naturali, etc. oppure stagionali o comunque periodici.

(c) A medio termine (da 1 a 5-10 anni). Il ruolo centrale è svolto dalle fasi di espansione e contrazione che coinvolgono più settori di attività

(d) A lungo termine (dai 10 anni in poi). Dominano i fattori demografici, le modifiche strutturali del sistema economico, l'evoluzione tecnologica, il sistema dei valori, etc.

La pianificazione non deve essere confusa con le previsioni. Un piano o programma di intervento è un insieme coordinato di azioni da intraprendere a fronte di uno o più scenari che potrebbero configurarsi. La previsione ha invece il compito di determinare il prevedibile corso degli eventi (unito ad una linea di sottostima ed una di sovrastima). Un piano di intervento richiede delle ottime previsioni (anche se talvolta si realizzano o debbono essere realizzati in assenza di previsioni), ma le previsioni possono ignorare i piani di intervento. Le previsioni solo il valore atteso, i piani sono il valore auspicato o temuto.

Sono disponibili metodi diversi per tempi diversi che hanno come denominatore comune nell'idea che i valori osservati siano privi di errori seri, che la rappresentazione univariata sia consona al problema trattato e che il fenomeno retrostante la serie storica si mantenga stabile per come è conosciuto ovvero le previsioni si effettuano

nel presupposto che non ci saranno modifiche strutturali drammatiche diverse da quelle già inglobate nella serie.

Per il medio-lungo termine la previsione non può limitarsi alla mera estensione del passato, ma comprendere varie alternative, anche accompagnate da considerazioni qualitative (scenari). Per il breve termine la previsione è legata all'inerzia intrinseca in ogni evoluzione dinamica che, peraltro, impedisce repentini cambiamenti. In questo caso le previsioni si estendono per più periodi, ma non troppi. Per il brevissimo termine la previsione è condizionata dalla situazione presente che si ritiene senz'altro continuata nell'immediato futuro. In genere basta traslare la previsione di un solo periodo.

C'è una "Terra di nessuno" che divide l'immaginazione creativa delle tecniche qualitative ed i modelli matematici delle tecniche quantitative che è quello delle diagnosi congiunturali. Analisi e conoscenza dell'evoluzione passata sono essenziali, ma più per individuare le relazioni fra variabili e regolarità di comportamento che non come fattore di predizione del futuro. In molte analisi, l'interesse si accentra su di un periodo tra i 12 e 24 mesi. Tale orizzonte è abbastanza ampio per dare spazio ad elementi decisionali, ma ancora breve perché si possa sentire l'influenza del passato più recente. Qui trovano largo impiego le tecniche BJ nonché le tecniche di decomposizione delle serie storiche e altri metodi di previsione.

Le tecniche di previsione falliscono quando il fenomeno rappresentato subisce spostamenti traumatici. In genere, il movimento è tale da portare i valori fuori dagli intervalli di previsione, anche quelli di più ampio raggio (scarsa precisione, ma elevata attendibilità). Occorre prudenza. Le previsioni servono spesso per decidere se intervenire o no su di un certo andamento. Intervenire al momento sbagliato può avere conseguente peggiori della omissione dell'intervento. D'altra parte non c'è neanche da sperare troppo nell'aiuto di variabili esplicative che accompagnino le previsioni univariate perché tendono a subire gli stessi mutamenti strutturali e spesso in contraddizione così da risultare poco utili. Non a caso, nelle competizioni sui metodi di previsioni, le tecniche che aggiungono come regressori delle variabili che spiegano il fenomeno, non sono tra le prime classificate.

5.4.2 Previsioni con la metodologia BJ

Le fasi precedenti della analisi BJ ci hanno convinto che il modello cui siamo pervenuti rappresenta in modo esauriente la serie e che, almeno per il momento, riteniamo ne possa riprodurre i comportamenti più tipici lasciando nei residui solo dei *white noise*. In questo senso è necessario che la serie storica sia stazionaria

in media perché la previsione realizzata nei modelli lineari è da intendersi come un valore atteso che deve essere legato alla struttura evolutiva del processo e non semplicemente alla posizione temporale in cui la previsione stessa è effettuata. E' poi necessaria la stazionarietà in varianza in quanto, alla previsione, si accompagna sempre un intervallo la cui ampiezza dipende proprio dallo scarto quadratico medio che deve essere fissa per essere utile e non soggetta a variazioni stocastiche nel corso del tempo (almeno che tali modifiche non siano inglobate nel metodo di stima).

La previsione con l'approccio BJ, nella versione da noi descritta, non può peraltro prevedere l'intervento di eventi che modificano sostanzialmente il comportamento della serie. Questi possono accadere anche nel brevissimo periodo. I cambiamenti bruschi possono essere inclusi nel processo solo se sono già avvenuti in passato adoperando delle variabili di comodo definite in corrispondenza dei periodo dei valori sospetti.

La estrapolazione delle componenti regolari e stagionali avviene con la traslazione temporale delle formule:

$$\hat{W}_{t+k|t} = E\left(W_{t+k}\,|W_1, W, 2, \cdots, W_n\right) \tag{5.34}$$

Quando i parametri sono noti si ha

$$a_t = W_t - \hat{W}_{t|t-1} \tag{5.35}$$

cioè le correzioni o innovazioni del processo sono sequenze di errori di previsione con un intervallo temporale pari ad un periodo. Ad esempio, un modello ARMA$(1,0,0)$ con $W_t = \mu + \phi_1 W_{t-1} + a_t$ può essere fatto scorrere in avanti con lo schema

$$W_{t+1} = \mu + \phi_1 W_t + a_{t+1} \tag{5.36}$$

Tuttavia, le correzioni/innovazioni a_t non possono essere osservate per periodi coevi o futuri per cui dobbiamo procedere alla previsione con il valore atteso della (5.36) condizionato alla conoscenza di W_t

$$\begin{aligned} E\left(W_{t+1}|W_t\right) &= \mu + \phi_1 W_t \\ E\left(W_{t+2}|W_t\right) &= \mu + \phi_1 E\left(W_{t+1}|W_t\right) = \mu + \phi_1\left(\mu + \phi_1 W_t\right) \end{aligned} \tag{5.37}$$

ed il processo può proseguire senza soluzione di continuità. La variabilità degli errori, intrinseca ad ogni processo stocastico, induce dispersione intorno al valore atteso e quindi ci aspettiamo di commettere un certo errore ad ogni previsione. Inoltre, dato lo schema di scorrimento mostrato nella relazione (5.37), gli errori si accumuleranno

per convergere ad livello predefinito qualore il processo fosse stazionario. Ad esempio nel modello ARMA$(1,0,0)$ si ha

$$\sigma^2\left(W_{t+j}\right) = \sigma_a^2\left[1+\phi_1^2+\phi_1^4+\phi_1^6+\cdots+\phi_1^{2(j-1)}\right] \tag{5.38}$$

che tende a $\sigma_a^2/\left(1-\phi_1^2\right)$. L'errore di previsione è tanto maggiore quanto più prossimo all'unità è il coefficiente ϕ_1 cioè quanto più simile alla non stazionarietà è il comportamento della serie. Infatti, l'intervallo di previsione è

$$\begin{aligned} \hat{W}_{t+1} \rightarrow \mu+\phi_1 W_t &\pm 1.96\sigma\left(W_{t+1}\right) \\ \hat{W}_{t+2} \rightarrow \mu\left(1+\phi_1\right)+\phi_1^2 W_t &\pm 1.96\sigma\left(W_{t+1}\right)\sqrt{\left(1+\phi_1^2\right)} \end{aligned} \tag{5.39}$$

L'algoritmo di estensione del modello ARMA procede allo stesso modo per ogni modello.

In generale, la funzione di previsione al tempo t è $\hat{W}_{t+j|t}$ vista come dipendente di k con k intero positivo. Nei capitoli precedenti si è potuto constatare che un modello ARIMA$(p,d,q)\times(P,D,Q)_s$ richiede la verifica di una serie di condizioni iniziali per passare poi alla soluzione delle equazioni ricorsive

$$\psi^*(B)\,\hat{W}_{t+k|t} = 0 \tag{5.40}$$

dove $\psi^*(B)$ è la versione auto-regressiva infinita del modello ARIMA $\psi^*(B) = \phi(B)\,\Phi\left(B^s\right)\Delta^d\Delta_s^D$ che include perciò delle radici unitarie, almeno quelle relative agli operatori differenze. Pertanto, la previsione può essere epressa con lo sviluppo convolutivo del polinomio $\psi^*(B)$

$$\begin{aligned} \hat{W}_{t+k|t} &= a_{t+k}+\psi_1 a_{t+k-1}+\cdots+\psi_{k-1}a_{t+1}+\psi_k a_t+\psi_{k+1}a_{t-1}+\cdots \\ &= \sum_{t=k}^{\infty}\psi_t a_{n+k-t} \end{aligned} \tag{5.41}$$

Poiché $E\left(a_{t+k}\right)=0$ e $E\left(a_{t-k}\right)=a_{t-k}$ per $k>0$, il valore atteso condizionato della (5.41) diventa

$$\hat{W}_{n+k|t} = E\left(W_{n+k|t}\right) = \sum_{t=0}^{\infty}\psi_{t+k}a_{t-k} \tag{5.42}$$

che esprime il valore previsto al periodo $n+k$ in base allo scenario noto al tempo n, con una combinazione lineare di correzioni/innovazioni presenti e passate. Lo scarto tra la (5.41) e la (5.46) definisce l'errore di previsione connesso come

$$e_{n+k|t} = W_{n+k}-\hat{W}_{n+k|t} = a_{n+k}+\psi_1 a_{n+k-1}+\cdots+\psi_{k-1}a_{n+1} \tag{5.43}$$

cioè un MA di ordine $k-1$, ma di correzioni future e quindi ignote. Il valore atteso degli errori è nullo per costruzione: $E\left[e_{n+k|t}\right]=0$. La varianza, in base alla formulazione BJ, è data da

$$\begin{aligned} var\left[e_{n+k|t}\right] &= var\left(a_{n+k}\right)+\psi_1^2 var\left(a_{n+k-1}\right)+\cdots+\psi_{k-1}^2 var\left(a_{n+1}\right) \quad (5.44)\\ &= \sigma_a^2\left(1+\psi_1^2+\cdots+\psi_{k-1}^2\right) \end{aligned}$$

Un primo problema da risolvere è che si dispone di una sere $W_t, t=1,2,\cdots,n$, ma non di una corrispondente serie per gli a_t. Per superare l'ostacolo, notiamo che $e_{n+1|t}=a_{n+1}$ (che è ancora incognito, visto che è un errore futuro, rispetto allo scenario di epoca n) cosicché

$$a_n = W_n - \hat{W}_{n|n-1} \quad (5.45)$$

Possiamo usare queste relazioni per calcolare a_k per ogni $k\leq n$, cioè calcoliamo la previsione per il ritardo uno e la deduciamo dal valore osservato. Questo va bene, procedendo all'indietro, fino a che rimaniamo nell'arco temporale della serie osservata. Poi capiterà di andare più indietro dell'inizio del periodo di osservazione. In questo caso, come abbiamo già fatto nella fase di stima del modello, potremmo porre $a_t=0$ cioè uguale alla media nulla nel presupposto che, trattandosi di valori lontani nel tempo rispetto all'epoca n, la scelta abbia poca rilevanza.

Un secondo problema è che non si conosce la varianza degli errori σ_a^2 del processo stocastixo lineare adottato come modello della serie storica. Anche in questo caso usiamo una procedura consolidata e cioè poniamo σ_a^2 uguale alla varianza degli errori stimati dal modello. $\hat{\sigma}_a^2$ (tutto funziona se il modello è quello giusto).

Un terzo problema è che si deve effettuare la somma di infiniti termini come richiesto alla relazione (5.41) che, in via di principio, è impossibile. Comunque i pesi, grazie alla stazionarietà del processo stocastico (con $d=D=0$) decadono abbastanza rapidamente verso lo zero e non si perde poi molto se la somma infinita diventa finita con il troncamento ad un certo termine. Se d e/o D non sono nulli, adotteremo la versione con singoli polinomi $\phi(B)\Phi(B^s)\Delta^d\Delta_s^D$ e $\theta(B)\Theta(B^s)$ e sforzandoci comunque di limitare il numero dei termini da considerare nella somma.

In breve, realizzare una previsione con l'approccio BJ è una procedura a tre passi. Innanzitutto si ricostruiscono le correzioni coeve al modello e le precedenti (ponendo queste pari a zero o altro valore ragionevole). Si usano poi queste insieme ai valori stimati della serie storica per ottenere $\hat{W}_{n+k|n}$. Le correzioni future a_{n+k} hanno pure, in media, valore zero; i valori futuri invece hanno valore atteso pari alle nostre previsioni (in base al modello prescelto) $\hat{W}_{n+k|n}, k=1,2,\cdots,$. La varianza stimata degli errori ci consente infine di fornire gli intervalli di confidenza delle previsioni.

Nella figura 5.18 è riportato l'esito delle previsioni del modello prescelto insieme agli intervalli di previsione (approssimati) per vari livelli confidenza, per dare un senso alla attendibilità del risultato.

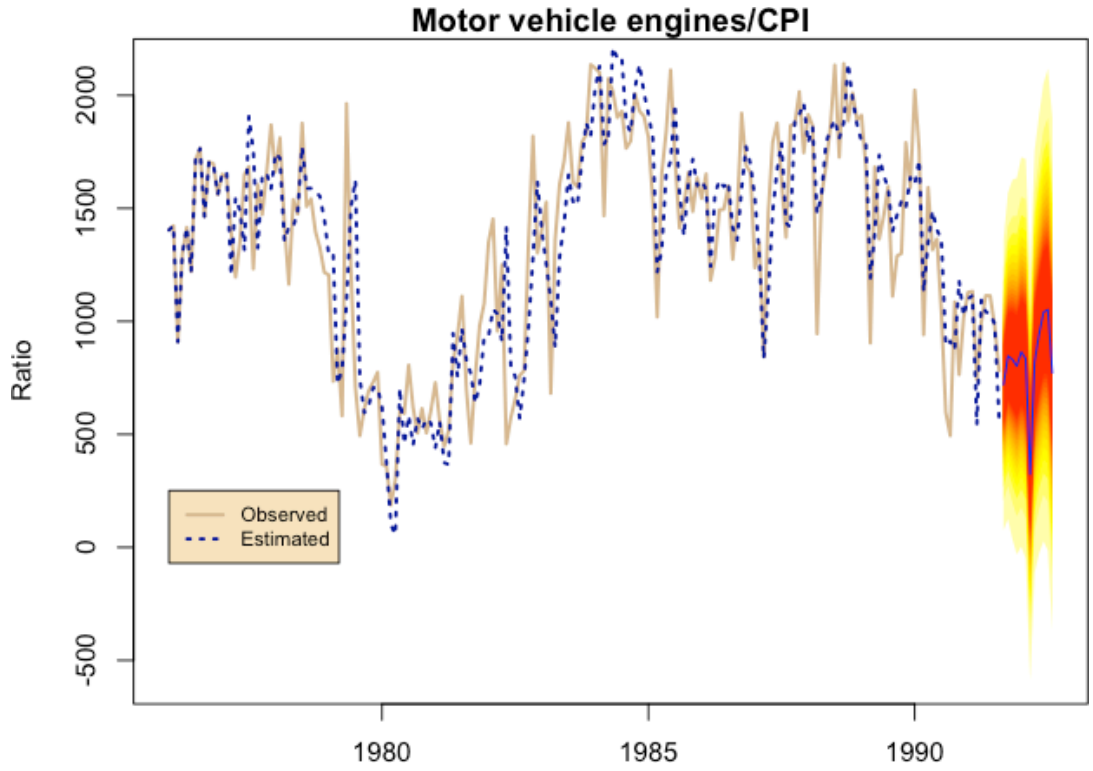


Figura 5.18: Previsioni con il modello ARIMA$(3,1,2) \times (0,1,1)_{12}$

5.4.3 Attendibilità delle previsioni (*cross-validation*)

Il successo di una tecnica di previsione si misura con l'errore medio che commette non tanto sui valori già usati per la stima perché a questo controllo hanno già provveduto le tecniche di selezione del modello ed abbiamo visto quanto possono essere simili sotto questo aspetto anche modelli diversissimi per struttura e genesi. Pertanto, per il controllo della qualità di modelli distinti è bene calcolare l'errore di previsione medio ottenuto accantonando ad esempio l'ultimo anno delle serie mensile ed utilizzandone i valori come termine di paragone (tecnica del *leaving-one-out*). Certo, la rinuncia a dati validi non si può fare a cuor leggero, ma il capitale di informazione è investito in una attività molto redditizia che ci conforterà sulla validità delle nostre scelte.

Tra le misure più ricorrenti per asseverare la bontà delle previsioni ottenuta in base da un certo modello debbono annoverarsi le seguenti:

$$\begin{aligned}
\text{RMSE (Relative Mean Squared Error)} &= \sqrt{\frac{1}{s}\sum_{t=n+1}^{n+s}\left(\frac{\hat{W}_t - W_t}{W_t+\epsilon}\right)^2} \qquad (5.46)\\
\text{MAPE (Mean Absolute Percentage Error)} &= \frac{1}{s}\sum_{t=n+1}^{n+s}\left|\frac{\hat{W}_t - W_t}{W_t+\epsilon}\right|\\
\text{XAPE (Max Absolute Percentage Error)} &= \max_{t\in\{n+1:n+s\}}\left|\frac{\hat{W}_t - W_t}{W_t+\epsilon}\right|\\
\text{Theil's } U_1 &= \frac{\sqrt{\sum_{t=n+1}^{n+s}\left(\hat{W}_t - W_t\right)^2}}{\sqrt{\sum_{t=n+1}^{n+s}\hat{W}_t^2}+\sqrt{\sum_{t=n+1}^{n+s}W_t^2}}\\
\text{Theil's } U_2 &= \sqrt{\frac{\sum_{t=n+1}^{n+s-1}\left(\frac{\hat{W}_{t+1} - W_{t+1}}{W_t}\right)^2}{\sum_{t=n+1}^{n+s-1}\left(\frac{W_{t+1} - W_t}{W_t}\right)^2}}
\end{aligned}$$

Le prime tre misure sono tre diverse metriche di Minkowski applicate agli errori relativi e pertanto la graduatoria dei modelli che ciascuna di esse fornisce è la stessa. La statistica U_1 di Theil varia tra zero ed uno; l'estremo inferiore indica perfetto adattamento e l'estremo superiore è raggiunto quando la successione dei valori previsti (o dei valori osservati) è nulla. La statistica U_2 è prossima all'unità se le previsioni realizzate con il modello sono analoghe a quelle ottenute con un metodo di previsione naive (proporzionale al valore osservato ne periodo precedente). Se la statistica è superiore ad uno il modello costituisce un miglioramento rispetto alla previsione naive e se maggiore di uno risulta addirittura in un peggioramento.

Il nostro software fornisce le statistiche (5.46) per i valori già osservati (*in-sample*). Nella figura 5.19 sono indicate quelle relative alla serie *Monthly motor vehicles engines - Canada*
Gli intervalli di previsione sono inadeguati vista la loro ampiezza e visto che propongono degli impossibili valori negativi. Basterebbe questo per richiedere un rifacimento *ex-novo* dell'intera procedura BJ.

La riflessione che sorge spontanea è la seguente: la rimozione, sia pure temporanea, di un certo numero di valori dalla serie storica (di quelli più recenti in partcolare), potrebbe forse portare ad identificare un modello diverso rispetto a quello che si otterrebbe impiegando l'intera serie di valori disponibile. In effetti,

	Point Forecast	True	%Err.	Lo 80	Hi 80	Lo 95	Hi 95
Sep 1990	787.44	494.54	59.23	237.00	1337.9	427.52	1147.3
Oct 1990	875.52	1085.67	19.36	147.57	1603.5	399.54	1351.5
Nov 1990	1139.67	764.47	49.08	303.09	1976.2	592.66	1686.7
Dec 1990	1066.77	1014.97	5.10	157.29	1976.2	472.09	1661.5
Jan 1991	1263.63	1128.26	12.00	303.17	2224.1	635.62	1891.7
Feb 1991	1134.75	1132.86	0.17	137.80	2131.7	482.88	1786.6
Mar 1991	820.30	741.25	10.66	-203.12	1843.7	151.13	1489.5
Apr 1991	1176.83	996.87	18.05	134.03	2219.6	494.98	1858.7
May 1991	1081.14	1114.03	2.95	24.05	2138.2	389.94	1772.3
Jun 1991	1142.09	1113.15	2.60	74.42	2209.8	443.98	1840.2
Jul 1991	1045.41	979.39	6.74	-30.11	2120.9	342.16	1748.7
Aug 1991	935.63	781.25	19.76	-145.73	2017.0	228.57	1642.7

Figura 5.19: Previsioni con il modello AR(13)

l'identificazione automatica della serie decurtata di un anno ha ritrovato il modello ARIMA$(3,1,2)\times(0,1,1)_{12}$ ancora primo tra i 7 migliori modelli. Non succede così per il modello auto-regressivo a grande ritardo perché la ricerca del miglior modello ha un esito nettamente diverso rispetto a quello dell'intera serie.

Le statistiche (5.46) sono state calcolate sia per ARIMA$(3,1,2)\times(0,1,1)_{12}$ che per ARIMA$(25,0,0)$ sulla base di un anno di osservazioni accantonate ottenendo

Tabella 5.1: *Confronto della bontà di adattamento per due modelli ARIMA*

Modello	RMSE	MAPE	XAPE
ARIMA $(3,1,2)\times(0,1,1)_{12}$	0.1337	0.0829	0.1514
AR $(25,0,0)$	0.4405	0.3083	0.8932

Come si può vedere, gli indicatori sono tutti a favore del modello da noi scelto.

CAPITOLO 6

RegArima

6.1 Processo Trend+ARMA

Nel 1983, Makridakis *et al.* avviarono la cosiddetta *M contest* in cui 111 serie storiche diverse furono analizzate a fini previsivi con le tecniche più in uso all'epoca della competizione. Si trattava di un innocente tentativo di fare il punto sullo stato dell'arte, ma ebbe un impatto straordinario sugli studi teorici delle serie storiche. Si accertò infatti che metodi di previone *naive* avevano avuto la meglio su processi con solidi impianti teorici e valide procedure di stima e algoritmi di calcolo molto efficaci. La gara è stata ripetuta nel 1993 (*M2 contest*) con 1101 serie storiche e da allora si sono susseguiti gli studi empirici per appurare il livello di attendibilità delle previsioni basate sulla metodologia Box-Jenkins.

Il confronto delle previsioni effettuate da modelli BJ con quelle effettuate con tecniche più antiche (quali ad esempio il metodo di scomposizione oppure il BV4) vide fin da subito l'approccio BJ eguagliato e superato nella qualità dei risultati. Dopo lustri di sperimentazioni ci si è convinti che, almeno nella versione più diffusa, la BJ non è adatta per fenomeni con forti componenti erratiche, spiccata eteroschedasticità e cambiamenti nelle tendenze anche a medio termine. Questi difetti divennero noti abbastanza presto e si è subito tentato di ovviare. Ma c'è anche un limite intrinseco: ad esempio, la trattazione del *trend* è poco duttile rispetto a quello che succede nei dati. Le semplici differenze, in scala naturale o logaritmica sono troppo drastiche per tenere conto di alcuni aspetti peculiari che si possono presentare nei dati.

Un approccio alternativo alla differenziazione è quello di non eliminare il *trend*, ma includerlo esplicitamente nel modello che rappresenta il processo. E' il cosiddetto approcio del modello *trend plus error* $Y_t = f(t) + W_t$ dove $f(t)$ è una funzione deterministica che dipende dall'ordinamento temporale t e W_t è un processo ARMA(p, q) stazionario. Spesso la $f(t)$ è espressa come un polinomio del tipo (3.77). La strategia qui accennata prevede come primo passo la regressione lineare multipla di Y_t sul polinomio in t o con un altra funzione linearizzabile quale la logistica, la esponenziale, la doppio logaritmica, etc. Successivamente, si effettua la modellazione ARMA(p, q) dei redisui $W_t = \left(Y_t - \hat{Y}_t\right)$. La debolezza maggiore di questa strategia consiste nel suo presupposto di una identica struttura valida su tutto l'arco temporale del processo (*trend* globale). Una assunzione in verità poco realistica soprattutto a fini previsivi. Nella figura 6.1 si delinea un *trend* quadratico fino al periodo t_1,

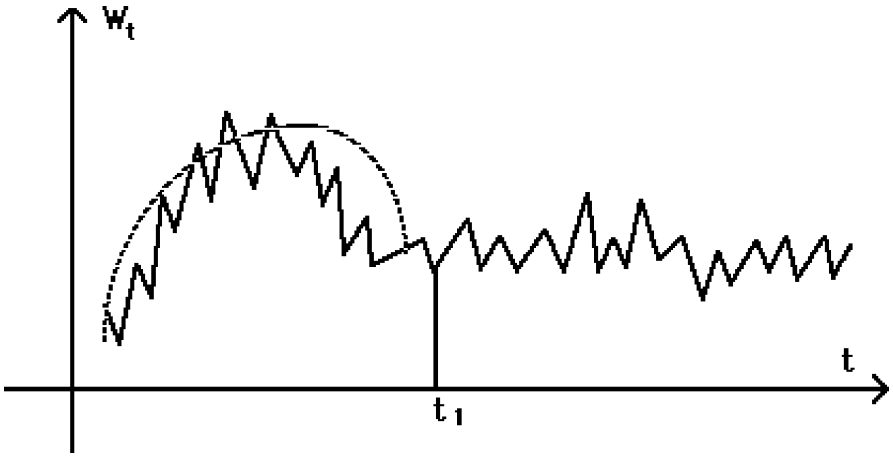


Figura 6.1: ARMA$(1, 1), \phi_1 = 0.7, \theta_1 = -0.5$

tuttavia la regressione di W_t su t e t_2 risentirebbe anche delle osservazioni dopo t_1 con una conseguenza doppiamente negativa: stima imprecisa del *trend* per il periodo fino a t_1 e previsioni future inefficaci (perchè il *trend* stimato si estenderebbe fino alla osservazione più recente).

Un metodo per superare il problema del trend che ha inclinazioni diverse in periodi diversi ci sarebbe. Il modello di regressione "broken stick" nel tempo con individuazione dei punti di frattura potrebbe rivelarsi molto utile in questo senso.