

Università degli studi di Genova
Facoltà di Scienze Matematiche, Fisiche e Naturali



Anno accademico 2003 – 2004

Corso di laurea in
Statistica Matematica e Trattamento Informatico dei Dati



Prova Finale

Titolo:

Analisi delle Serie Storiche e Applicazioni

Relatori:
Dott. Fabio Rapallo
Prof. Mauro Gasparini

Correlatore:
Dott. Emanuela Sasso

Candidato:
Daniele Rampoldi

Indice

Prefazione	5
1 Processi Stocastici e Serie Storiche	7
1.1 Correlazione e indipendenza	8
1.2 Distribuzioni di probabilità	9
1.3 Stazionarietà e autocovarianza	9
1.4 Stazionarietà per processi Gaussiani	11
1.5 Processi invertibili e periodici	11
1.6 La funzione di autocorrelazione	12
1.7 Un esempio di funzione di autocorrelazione: processo WN	16
1.8 Ergodicità e stima delle funzioni di autocorrelazione	16
1.9 La funzione di autocorrelazione parziale	20
2 La classe dei modelli ARIMA	23
2.1 La genesi dei modelli ARMA	23
2.1.1 Il Teorema di Wold	24
2.1.2 Il modello Media Mobile (MA)	25
2.1.3 Invertibilità nei modelli Media Mobile	26
2.1.4 L'operatore backward e il modello Auto Regressivo (AR)	27
2.1.5 I modelli misti (ARMA)	29
2.2 Serie evolutive e modelli ARIMA	31
3 La costruzione di un modello ARIMA	35
3.1 Identificabilità dei modelli ARMA	35
3.2 Costruzione iterativa di un modello ARIMA	36
3.2.1 Analisi preliminari	37
3.2.2 L'identificazione del modello	39
3.2.3 La stima dei parametri	40
3.2.4 La verifica del modello	41

Indice

4	Applicazioni a dati economici	43
4.1	Descrizione del problema	43
4.2	Descrizione della procedura ARIMA	43
4.2.1	Lo statement identify	44
4.2.2	Lo statement estimate	44
4.2.3	Lo statement forecast	45
4.3	Dati analizzati	46
4.4	Analisi del prezzo del petrolio al barile	46
4.5	Analisi della produzione di elettricità in Australia	52
	Bibliografia	59
	Ringraziamenti	61

Prefazione

Cosa si intende per serie storica

Per “Serie Storica” o “Serie Temporale” intendiamo una successione di osservazioni ordinate logicamente secondo una variabile t , che solitamente rappresenta il tempo.

Per tale motivo si cerca di descrivere, modellare e quindi prevedere la serie storica nella sua dinamica temporale (analisi univariata) e nelle sue connessioni con altre serie storiche ad essa collegate (analisi multivariata).

L’analisi classica

All’inizio del ventesimo secolo, in seguito all’esigenza di descrivere il comportamento delle serie storiche con semplici funzioni del tempo che esprimessero il risultato finale mediante “componenti” prefissate, nacque l’analisi classica delle serie storiche.

Questo tipo di studio si basa sul considerare la variabile casuale X_t come la somma di una funzione deterministica $f(t)$, che rappresenta l’evoluzione di fondo (trend) del fenomeno, di una componente ciclica $c(t)$, che identifica un ciclo periodico dovuto a cause strutturali (come per esempio un particolare ciclo economico), di una componente stagionale $s(t)$, di periodo esattamente legato al ciclo solare annuale, e di una componente accidentale A_t , che rappresenta il residuo non spiegato dal modello e che costituisce quindi la componente aleatoria di X_t .

La decomposizione così ottenuta:

$$X_t = f(t) + c(t) + s(t) + A_t$$

ha il pregio di una grande semplicità sia nelle ipotesi che nella metodologia utilizzata per dividere le componenti.

Accanto a questo pregio vi sono però alcuni difetti, il più criticato dei quali è che la scissione e il metodo statistico utilizzato sono “dati una volta per tutte”, senza

Prefazione

considerare la struttura probabilistica della serie storica e le eventuali informazioni a disposizione del ricercatore.

I modelli di Box e Jenkins

Cercando di rimediare ai difetti dell'analisi classica Box e Jenkins proposero degli strumenti statistici affinché le serie "speak for themselves", ossia cercarono di costruire un approccio ai dati per il quale sia la serie storica ad orientare verso il modello e non viceversa.

Questo permette un approccio globale per l'analisi statistica di una classe molto ampia di serie storiche, permettendoci dunque di rispondere ed affrontare una grande varietà di problemi, anche se ciò non implica che tale metodo ottenga una risposta ottimale per tutti i casi affrontati.

Capitolo 1

Processi Stocastici e Serie Storiche

Un processo stocastico X_t è una famiglia di variabili casuali descritte da un parametro t appartenente ad un insieme T .

La conoscenza di un processo stocastico equivale dunque alla conoscenza della distribuzione di probabilità multivariata per qualsiasi sottoinsieme di T .

Poiché nei problemi di cui ci occuperemo dovremo trattare fenomeni rilevati ad intervalli equispaziati, possiamo considerare d'ora in avanti le successioni di variabili casuali continue X_t , $t \in T$, con $T = \{1, 2, 3, \dots\}$, cioè i processi stocastici continui a parametro discreto.

Un tale processo è quindi la famiglia di variabili casuali $\{X_1, X_2, \dots\}$ per la cui conoscenza occorre specificare la funzione di distribuzione di ciascuna k -upla $(X_{t_1}, \dots, X_{t_k})$ di variabili casuali, per ogni k e per ogni k -upla (t_1, \dots, t_k) di indici.

Una serie storica è una parte finita di una singola realizzazione del processo; quindi non solo costituisce un campione unico della famiglia di variabili casuali che caratterizzano il processo, ma è anche un campione troncato perchè si osserva solo per $t = 1, \dots, N$.

D'altro canto, dato un processo X_t , esistono infinite possibili realizzazioni, tra le quali noi osserviamo unicamente la successione dei risultati campionari (x_1, \dots, x_N) , chiamata anche realizzazione o traiettoria del processo.

Questi problemi ci costringono dunque ad una limitazione dei processi stocastici considerati, perchè solo in una classe più ristretta sarà possibile dedurre informazioni consistenti dalla realizzazione finita osservata.

Processi Gaussiani, stazionari e invertibili

È possibile stabilire, sulla base delle variabili casuali componenti un processo e dei loro legami, una classificazione dei processi stocastici.

1.1 - Correlazione e indipendenza

Una prima distinzione tra i differenti processi stocastici riguarda l'indipendenza o meno delle variabili casuali che lo compongono.

La quasi totalità dei processi che vedremo sarà a componenti correlate, quindi non indipendenti.

L'unica eccezione alla non indipendenza sarà il processo definito "White Noise" (Rumore Bianco) di valor medio nullo e di varianza costante σ_a^2 (quindi non dipendente da t), che indicheremo:

$$A_t \sim WN(0, \sigma_a^2).$$

Pertanto un processo stocastico WN è caratterizzato da:

$$E(A_t) = 0;$$

$$E(A_t^2) = \text{Var}(A_t) = \sigma_a^2;$$

$$\text{Cov}(A_t, A_s) = E(A_t A_s) = 0, \quad \forall t \neq s.$$

In generale non si fanno ipotesi a priori sulle variabili casuali A_t , ma, se si suppone che, per ogni t , A_t sia anche Normale, si parla allora di processo "White Noise Gaussiano".

Poiché la non correlazione di variabili casuali Normali implica l'indipendenza, un processo White Noise Gaussiano è a componenti indipendenti.

1.2 - Distribuzioni di probabilità

Una seconda distinzione riguarda la distribuzione di probabilità delle variabili casuali componenti il processo.

L'ipotesi più comune è quella di supporre che le variabili casuali $(X_{t_1}, \dots, X_{t_k})$ abbiano legge normale multivariata per ogni (t_1, \dots, t_k) e per ogni $k \geq 1$.

In tal caso il processo stocastico X_t si definisce processo Gaussiano e possiede funzione di densità multivariata:

$$\varphi_{1, \dots, k}(x_{t_1}, \dots, x_{t_k}) = (2\pi)^{-k/2} |\Sigma|^{1/2} \exp\left[-1/2(x_t - \mu_t)' \Sigma^{-1}(x_t - \mu_t)\right],$$

essendo $\mu_t = (E(X_t))$ il vettore dei valori medi e Σ la matrice di dimensione $k \times k$ il cui elemento generico è $\Sigma_{ij} = (Cov(X_{t_i}, X_{t_j}))$. È interessante notare che un processo Gaussiano è caratterizzato solo da μ_t e Σ .

Questa osservazione è di particolare rilievo perchè afferma che in una certa classe di processi stocastici (nel nostro caso i processi Gaussiani) la conoscenza del processo stocastico (e quindi la conoscenza di tutte le funzioni del processo) può essere ricondotta alla conoscenza di una particolare categoria di funzioni (i momenti misti) a loro volta stimabili dalle realizzazioni finite (e quindi dalle serie storiche in nostro possesso).

1.3 – Stazionarietà e autocovarianza

Si possono fare altre distinzioni considerando il comportamento della famiglia di variabili casuali rispetto alla variabile temporale.

Si può ipotizzare, per esempio, che la funzione di distribuzione multivariata sia indipendente da t o che alcuni momenti caratteristici siano indipendenti da t . Questo non implica che tali processi e le loro realizzazioni siano costanti, ma significa che la distribuzione di probabilità (nel primo caso) o alcuni valori medi (nel secondo caso) presentano una formulazione analitica che non dipende da t .

I processi stocastici con struttura probabilistica o con valori medi invarianti rispetto a t vengono definiti stazionari.

Più precisamente un processo X_t è stazionario in senso stretto (o forte) se la distribuzione multivariata delle variabili casuali $(X_{t_1}, \dots, X_{t_k})$ non è funzione di (t_1, \dots, t_k) , per ogni $k \geq 1$. Se esistono, tutti i momenti di un processo stazionario non sono funzioni di t .

Un processo X_t è stazionario in senso debole (o stazionario in covarianza o stazionario del secondo ordine) se valgono le seguenti condizioni:

- (1) $E(X_t) = \mu, \quad \forall t$
- (2) $E(X_t - \mu)^2 = \sigma^2 < +\infty, \quad \forall t$
- (3) $E((X_t - \mu)(X_s - \mu)) = \gamma(|s - t|), \quad \forall \text{ coppia } (t, s)$

La prima condizione richiede semplicemente che il valor medio del processo X_t sia costante e pari a μ al variare di t .

La seconda condizione, in maniera analoga, richiede che la varianza del processo (pari a σ^2) sia finita e costante al variare di t .

L'ultima condizione implica che per ogni t ed s esista una funzione di $|s - t|$, detta autocovarianza (o funzione di autocovarianza) e indicata con $\gamma(\cdot)$, che misura la covarianza fra le variabili casuali X_t e X_s ; in tale ipotesi la covarianza dipende solo dalla "distanza nel tempo" fra le variabili.

Un processo che soddisfa le tre condizioni precedenti viene anche detto processo stazionario.

È da notare come la stazionarietà in senso forte implichi la stazionarietà in senso debole, mentre non vale il contrario.

Osserviamo inoltre che non è restrittivo considerare processi stazionari Z_t a valor medio nullo, perchè è sufficiente porre $Z_t = X_t - E(X_t)$. Per processi stazionari a valor medio nullo la varianza vale $\sigma^2 = E(Z_t^2)$ e coincide con il momento secondo, mentre l'autocovarianza $\gamma(\cdot)$ diviene, posto $s = t + k$:

$$\gamma(|t - s|) = \gamma(k) = \text{Cov}(Z_t, Z_{t+k}) = E(Z_t Z_{t+k}) = E(Z_t Z_{t-k}).$$

La condizione (3) implica dunque che per un processo stazionario la covarianza tra le variabili casuali distanti k intervalli temporali (*lags* nel gergo inglese) è funzione del lag k , ma non di t .

L'autocovarianza $\gamma(k)$ misura, quindi, il legame lineare fra Z_t e Z_{t+k} al variare di k .

La stazionarietà impone che la covarianza fra Z_t e Z_{t+k} , Z_{t+1} e Z_{t+1+k} , e così via, sia costante, anche se sarà in generale diversa dalla covarianza fra Z_t e Z_{t+k+1} , Z_t e Z_{t+k+2} .

Osserviamo infine che $\gamma(0) = E(Z_t^2) = \text{Var}(Z_t)$ non dipende da t . Quindi un processo stazionario di valor medio nullo è un processo stocastico con varianza finita e autocovarianza che dipende solo dal lag fra le variabili casuali Z_t e Z_{t+k} .

Riassumendo possiamo dunque affermare che per verificare se un processo stocastico X_t è stazionario, accertato che il valor medio di X_t non dipende da t , basta verificare che $\gamma(0)$ (cioè $\text{Var}(X_t)$) sia finita e $\gamma(k)$, per $k = 1, 2, \dots$, sia una funzione di k , ma non di t .

1.4 – Stazionarietà per processi Gaussiani

Se un processo Gaussiano è stazionario in senso debole, allora il suo vettore dei valori medi non è funzione di t , né lo è la matrice Σ delle autocovarianze.

Quindi un processo Gaussiano stazionario in senso debole possiede funzione di densità multivariata che non è funzione di t , il che implica che è anche stazionario in senso forte.

Pertanto, per processi Gaussiani, la stazionarietà in senso forte e quella in senso debole coincidono.

1.5 – Processi invertibili e periodici

Un processo stocastico X_t è invertibile se esiste una funzione lineare $h(\cdot)$ e un processo WN A_t tale che, per ogni t , si possa scrivere:

$$X_t = h(X_{t-1}, X_{t-2}, \dots) + A_t .$$

L'invertibilità è quindi la possibilità di esprimere un processo X_t tramite le variabili casuali del "passato". A tale relazione viene aggiunto il processo WN perchè, in assenza di esso, X_t sarebbe una funzione deterministica di t .

Diremo inoltre che un processo X_t è periodico se esiste un valore s tale che, per ogni t , $X_t = X_{t-s}$ con probabilità 1. Il valore s è detto periodo.

Se s è esattamente pari all'anno solare, diremo che il processo periodico è stagionale.

1.6 – La funzione di autocorrelazione

Grazie alla funzione di autocovarianza possiamo verificare se un processo X_t è stazionario.

L'autocovarianza, come covarianza fra X_t e X_{t+k} misura il segno e la forza del legame lineare esistente fra X_t e X_{t+k} al variare di k .

Tuttavia, proprio perchè $\gamma(k)$ è una covarianza, non è agevole effettuare confronti fra differenti processi stocastici, perchè la funzione varia in tutto il campo dei reali ed è dipendente da un cambiamento di scala del processo.

In analogia con il coefficiente di correlazione si introduce quindi la funzione di autocorrelazione, indicata con $\rho(k)$ e definita come il coefficiente di correlazione lineare fra le variabili casuali X_t e X_{t+k} , al variare di k .

Quindi:

$$\rho(k) = E\left[\left(\frac{X_t - \mu}{\sigma}\right)\left(\frac{X_{t+k} - \mu}{\sigma}\right)\right] = \frac{\text{Cov}(X_t, X_{t+k})}{\sqrt{\text{Var}(X_t)\text{Var}(X_{t+k})}}, \text{ per } k = 0, 1, 2, \dots$$

Poiché $\text{Var}(X_t) = \text{Var}(X_{t+k}) = \gamma(0)$, la definizione si semplifica in:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)}, \text{ per } k = 0, 1, 2, \dots$$

La funzione di autocorrelazione $\rho(k)$, chiamata anche autocorrelazione o funzione di autocorrelazione globale, possiede notevoli proprietà derivate da quelle del coefficiente di correlazione.

Fra di esse le più importanti sono:

$$(1) \rho(0) = 1$$

Infatti $\rho(0) = \gamma(0)/\gamma(0)$ ed inoltre $\rho(0)$ è il coefficiente di correlazione fra X_t e se stesso.

$$(2) \rho(k) = \rho(-k) \text{ per ogni } k = 0, 1, 2, \dots$$

Infatti $\gamma(k) = \gamma(-k)$ perchè

$$\gamma(k) = \text{Cov}(X_t, X_{t+k}) = \text{Cov}(X_{t-k}, X_t) = \text{Cov}(X_t, X_{t-k})$$

per la simmetria della covarianza.

$$(3) |\rho(k)| \leq 1, \text{ per } k = 0, 1, 2, \dots$$

Infatti per la disuguaglianza di Schwartz $\gamma^2(k) \leq \gamma^2(0)$ perchè

$$[\text{Cov}(X_t, X_{t+k})]^2 \leq \text{Var}(X_t)\text{Var}(X_{t+k})$$

e dalla stazionarietà del processo segue la proprietà.

$$(4) \rho_{aX_t+b}(k) = \rho_{X_t}(k), \text{ per } k = 0, 1, 2, \dots, \text{ e per qualsiasi coppia reale } (a, b), a \neq 0.$$

Infatti, essendo $\text{Cov}(aX_t + b, aX_{t+k} + b) = a^2 \text{Cov}(X_t, X_{t+k})$, poiché

$$\text{Var}(aX_t + b) = a^2 \text{Var}(X_t) \text{ segue:}$$

$$\text{Cov}(aX_t + b, aX_{t+k} + b) / \text{Var}(aX_t + b) = \text{Cov}(X_t, X_{t+k}) / \text{Var}(X_t) = \rho_{X_t}(k).$$

Ricordiamo a questo punto che la matrice di Toeplitz di ordine m , associata a $\rho(k)$, è definita per ogni $m = 1, 2, \dots$, da:

$$\mathbf{P}_{(m)} = \begin{pmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(m-1) \\ \rho(1) & 1 & \rho(1) & \dots & \rho(m-2) \\ \rho(2) & \rho(1) & 1 & \dots & \rho(m-3) \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \rho(m-1) & \rho(m-2) & \rho(m-3) & \dots & 1 \end{pmatrix} = \{\rho(|i-j|)\}, \text{ per } (i, j) = 1, \dots, m$$

Tale matrice è doppiamente simmetrica perchè è simmetrica e tutte le diagonali possiedono elementi comuni.

Questa particolare matrice ci è utile per definire una nuova proprietà:

- (5) La matrice di Toeplitz di ordine m associata alla funzione di autocorrelazione di un processo stazionario è definita positiva.

Un'ultima proprietà importante, che però riguarda un ambito delle serie storiche che noi non studieremo (l'analisi spettrale), è la

- (5 bis) Se $\rho(k)$ è una funzione di autocorrelazione, esiste allora una funzione

non negativa $\psi(\omega)$, per $-\pi < \omega < \pi$, tale che $1 = \int_{-\pi}^{\pi} \psi(\omega) d\omega$,

e per la quale si avrà:

$$\psi(\omega) = \frac{1}{2\pi} \left[1 + 2 \sum_{k=1}^{\infty} \rho(k) \cos(\omega k) \right].$$

La funzione $\psi(\omega)$ viene chiamata funzione di densità spettrale del processo stazionario X_t .

Commentiamo ora brevemente le proprietà enunciate.

La (1) e la (3) ci informano che $\rho(k)$ è normalizzata a 1 ed ha massimo per $k = 0$.

La (2) mostra che la funzione di autocorrelazione è simmetrica rispetto a $k = 0$: l'analisi sarà quindi esplicitata solamente per $k > 0$.

La (4) mostra che l'autocorrelazione è invariante per traslazione e cambiamento di scala.

La (5) e la sua equivalente (5 bis), infine, esprimono una serie di condizioni necessarie e sufficienti perchè una successione $\{1, \rho(1), \rho(2), \dots, \rho(m), \dots\}$ sia effettivamente la funzione di autocorrelazione di un processo stazionario.

In particolare la (5) esprime condizioni in sequenza sui $\rho(k)$: innanzi tutto ricordiamo che una matrice è definita positiva se e solo se tutti i minori principali di qualsiasi ordine sono positivi.

Questo, per il nostro caso, implica che:

$$|\mathbf{P}_{(1)}| = |1| > 0 ;$$

$$|\mathbf{P}_{(2)}| = \begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{vmatrix} > 0 \rightarrow 1 - \rho^2(1) > 0 \rightarrow -1 < \rho(1) < +1 ;$$

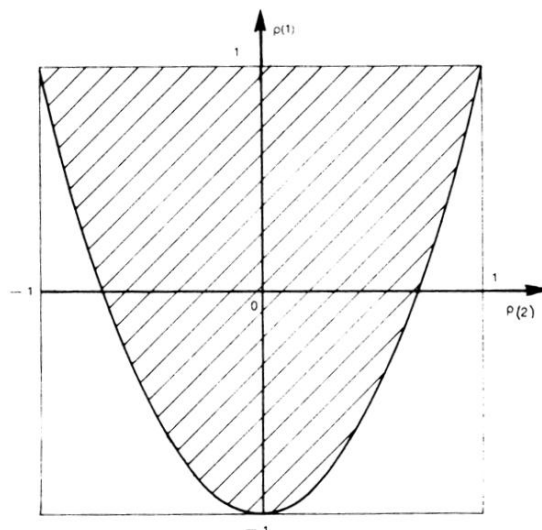
$$|\mathbf{P}_{(3)}| = \begin{vmatrix} 1 & \rho(1) & \rho(2) \\ \rho(1) & 1 & \rho(1) \\ \rho(2) & \rho(1) & 1 \end{vmatrix} > 0 \rightarrow [1 - \rho(2)][1 + \rho(2) - 2\rho^2(1)] > 0 \rightarrow$$

$$\rightarrow \begin{cases} -1 < \rho(2) < +1 \\ \rho(2) > 2\rho^2(1) - 1 \end{cases}$$

e così via per $|\mathbf{P}_{(m)}| > 0$, per $m = 4, 5, \dots$

Questa serie di vincoli riduce “a cascata” lo spazio di definizione teorico dei $\rho(k)$: così $\rho(2)$ sarà vincolato da $\rho(1)$, $\rho(3)$ sarà vincolato da $\rho(2)$ e $\rho(1)$, e così via. Nella figura riportata sotto è visualizzato lo spazio di definizione per le prime due autocorrelazioni di un processo stazionario.

In generale $\rho(m)$ è vincolato a tutti i $\rho(k)$, per $k < m$.



1.7 – Un esempio di funzione di autocorrelazione: processo WN

Come abbiamo visto in precedenza una classe di processi molto interessanti nell'ambito dei processi stocastici sono i processi White Noise.

La funzione di autocorrelazione del processo $A_t \sim WN(0, \sigma_a^2)$ è particolarmente semplice. Infatti, essendo $E(A_t A_s) = 0$ per $t \neq s$, sarà:

$$\gamma(k) = \begin{cases} \sigma_a^2, & \text{per } k = 0 \\ 0, & \text{per } k \neq 0 \end{cases},$$

il che implica

$$\rho(k) = \begin{cases} 1, & \text{per } k = 0 \\ 0, & \text{per } k \neq 0 \end{cases}.$$

1.8 – Ergodicità e stima della funzione di autocorrelazione

Poiché la funzione di autocorrelazione $\rho(k)$ è una misura della struttura interna del processo stazionario X_t , assume particolare interesse la sua stima statistica a partire da una realizzazione finita, cioè la serie storica $\{x_t, t = 1, \dots, N\}$.

Il problema si complica perchè la conoscenza di (x_1, \dots, x_N) è una informazione congiunta sulle variabili casuali (X_1, \dots, X_N) e non sulla generica variabile casuale X_t che definisce (al variare di t) il processo.

Quindi la serie storica è una successione ordinata di N campioni di dimensione unitaria su N distinte variabili casuali, generalmente fra di loro non indipendenti e non somiglianti. Questa complicazione è superata grazie al concetto di ergodicità: “un processo stocastico X_t è ergodico rispetto ad un parametro φ se la stima temporale del parametro $\hat{\varphi}(X_1, \dots, X_T)$, ottenuta da una serie storica, converge in media quadratica a quel parametro”, ossia se $\lim_{T \rightarrow \infty} E(\hat{\varphi}(X_1, \dots, X_T) - \varphi)^2 = 0$.

È dunque molto importante stabilire condizioni necessarie e sufficienti perchè i processi siano ergodici.

Una delle condizioni più comuni è la seguente: condizione necessaria e sufficiente affinché il processo sia ergodico rispetto al valor medio μ è che la sua funzione di autocorrelazione tenda a zero.

È quindi molto importante sottolineare come solo per processi ergodici ci si può porre correttamente il problema dell'inferenza statistica sulle serie storiche. Infatti l'ergodicità garantisce che dall'unica informazione disponibile (la serie storica) si possa risalire a stime consistenti del processo stazionario X_t .

Dopo questa importante premessa, descrivendo gli stimatori a cui siamo interessati, ci riferiremo sempre a processi ergodici.

La stima dell'autocorrelazione $\rho(k) = \gamma(k)/\gamma(0)$ avviene in maniera piuttosto semplice, costruendo stimatori per $\gamma(k)$, $k = 0, 1, 2, \dots$, e, di conseguenza, costruendo lo stimatore $\hat{\rho}(k) = \hat{\gamma}(k)/\hat{\gamma}(0)$, $k = 1, 2, \dots$.

Il problema di stimare l'autocorrelazione si traduce quindi in un problema di stima di autocovarianza.

Innanzitutto il valor medio $\mu = E(X_t)$ viene stimato mediante $\bar{X} = \sum_{t=1}^N X_t / N$ che è non distorto, consistente e asintoticamente Normale, e la cui varianza vale:

$$\text{Var}(\bar{X}) = \frac{1}{N} \left[\gamma(0) + 2 \sum_{j=1}^{N-1} \left(1 - \frac{j}{N}\right) \gamma(j) \right] = \frac{\gamma(0)}{N} \left[1 + 2 \sum_{j=1}^{N-1} \left(1 - \frac{j}{N}\right) (\rho(j)) \right].$$

In pratica, a causa della correlazione esistente fra X_t e X_{t-k} , la varianza di X (che per campioni indipendenti vale σ^2/N) viene alterata dal fattore in parentesi quadra.

Dopo aver stimato μ con \bar{X} , conviene considerare il processo scarto $Z_t = X_t - \mu$ che, essendo un semplice traslazione di X_t , possiede media zero e la stessa varianza, autocovarianza e autocorrelazione di X_t .

Infine, per quel che riguarda la stima dell'autocovarianza, sono stati proposti diversi stimatori. Fra i più utilizzati vi sono:

$$\hat{\gamma}_1(k) = \frac{1}{N} \sum_{t=1}^N Z_t Z_{t+k}, \text{ purché si ponga } Z_{t+N} = Z_t \text{ per } t \geq 1;$$

$$\hat{\gamma}_2(k) = \frac{1}{N-k} \sum_{t=1}^{N-k} Z_t Z_{t+k};$$

$$\hat{\gamma}_3(k) = \frac{1}{N} \sum_{t=1}^{N-k} Z_t Z_{t+k} .$$

Il primo stimatore $\hat{\gamma}_1(k)$ possiede notevoli proprietà teoriche, ma richiede l'esistenza di un universo "circolare" ($Z_{t+N} = Z_t$), ovvero un processo che si ripete dopo N unità temporali.

Per N elevato si nota poi che $\hat{\gamma}_2(k) \approx \hat{\gamma}_3(k)$, con una differenza piuttosto marcata solo per piccoli campioni.

Inoltre è stato dimostrato che $\hat{\gamma}_2(k)$ è uno stimatore corretto per $\gamma(k)$, mentre $\hat{\gamma}_3(k)$, pur essendo distorto, fornisce un errore quadratico medio inferiore a $\hat{\gamma}_2(k)$ (quindi $\hat{\gamma}_3(k)$ è più efficiente di $\hat{\gamma}_2(k)$), ma soprattutto soddisfa la proprietà per cui $\hat{\mathbf{P}}$ è una matrice (stimata) definita positiva, proprio come \mathbf{P} , il che non avviene necessariamente costruendo $\hat{\mathbf{P}}$ con lo stimatore $\hat{\gamma}_2(k)$.

Per questi motivi e per la sua semplicità, si preferisce lo stimatore $\hat{\gamma}_3(k)$.

La stima della funzione di autocorrelazione avviene quindi generalmente tramite:

$$\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)} = \frac{\sum_{t=1}^{N-k} Z_t Z_{t+k}}{\sum_{t=1}^N Z_t^2} = \frac{\sum_{t=1}^{N-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^N (X_t - \bar{X})^2} , \quad k = 0, 1, 2, \dots$$

Naturalmente lo stimatore $\hat{\rho}(k)$ è tale che $\hat{\rho}(-k) = \hat{\rho}(k)$, per cui viene stimato solo per k positivo.

La distribuzione di $\hat{\rho}(k)$ ha creato nel passato grossi problemi statistici, ma adesso si tende ad utilizzare formule approssimate, spesso piuttosto semplici. Fra le più semplici ricordiamo:

$$\text{Var}[\hat{\rho}(k)] \approx 1/N , \text{ per ogni } k = 1, 2, \dots ,$$

che, soprattutto per lags non troppo bassi, può ritenersi accettabile.

Tale varianza asintotica è dedotta sotto l'ipotesi $H_0 : X_t \sim WN$. Quindi, costruendo gli intervalli $(\hat{\rho}(k) \pm z_{\alpha/2} / \sqrt{N})$, si possono ottenere bande di confidenza approssimate e quindi decidere sulla nullità di $\rho(k)$ per mezzo di un test statistico.

Una migliore approssimazione si può ottenere con la formula:

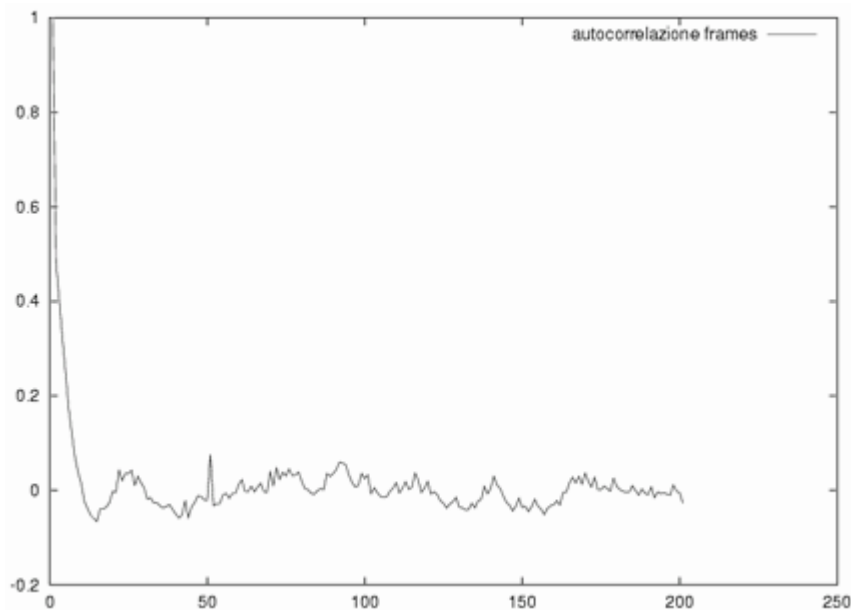
$$\text{Var}(\hat{\rho}(k)) \approx \frac{1}{N} \left\{ 1 + 2 \sum_{k=1}^{\infty} [\rho(k)]^2 \right\}$$

che però richiede ipotesi più specifiche e calcoli più complessi.

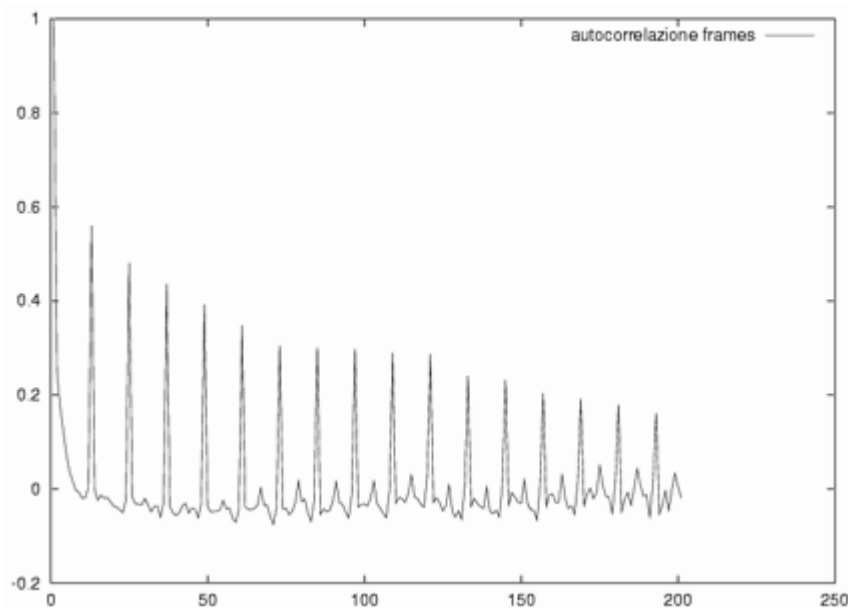
Per calcolare queste stime, partendo dalla serie storica (x_1, \dots, x_N) , si calcolano innanzi tutto la serie degli scarti $z_i = x_i - \bar{x}$, con $i = 1, \dots, N$ (con $\bar{x} = \sum x_i / N$).

Quindi si calcola la stima di $\gamma(k)$ come descritto sopra, fino ad un massimo lag che l'esperienza ha posto pari a $N/4$ o al suo intero successivo, e, infine, si calcola $\hat{\rho}(k)$ dalle stime precedenti.

È infine interessante il correlogramma, che è la rappresentazione grafica e dunque più immediata della stima delle funzioni di autocorrelazione trovate sopra. Proponiamo qui sotto due esempi di correlogrammi piuttosto interessanti e caratteristici.



La figura sopra rappresenta una funzione di autocorrelazione piuttosto “classica”: la funzione ha un massimo per $k = 0$ ($\rho(0) = 1$), ma scende piuttosto rapidamente, fino ad oscillare attorno allo zero, valore a cui dovrebbe tendere per $k \rightarrow \infty$.



La seconda figura rappresenta invece una funzione di autocorrelazione corrispondente ad un processo stagionale. Si nota infatti un picco (che diminuisce sempre più con il passare del tempo) ad ogni lag di dodici mesi che indica una elevata correlazione tra i dati di un certo periodo dell'anno.

1.9 – La funzione di autocorrelazione parziale

Accanto alla funzione di autocorrelazione globale $\rho(k)$, si introduce, nell'analisi dei processi stazionari, la funzione di autocorrelazione parziale.

Questa, generalmente indicata con $\pi(k)$, è definita al lag k , per $k = 0, \pm 1, \pm 2, \dots$, come la correlazione fra Z_t e Z_{t+k} al netto della correlazione esistente fra le variabili casuali "intermedie" (cioè $Z_{t+1}, \dots, Z_{t+k-1}$).

La funzione di autocorrelazione parziale è data dal rapporto fra due determinanti:

$$\pi(k) = \frac{|\mathbf{P}_{(k)}^*|}{|\mathbf{P}_{(k)}|}, \text{ per } k = 0, 1, 2, \dots,$$

con:

$$\mathbf{P}_{(k)} = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(k-1) \\ \rho(1) & 1 & \rho(1) & \dots & \rho(k-2) \\ \rho(2) & \rho(1) & 1 & \dots & \rho(k-3) \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \rho(k-1) & \rho(k-2) & \rho(k-3) & \dots & 1 \end{bmatrix};$$

$$\mathbf{P}_{(k)}^* = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(1) \\ \rho(1) & 1 & \rho(1) & \dots & \rho(2) \\ \rho(2) & \rho(1) & 1 & \dots & \rho(3) \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \rho(k-1) & \rho(k-2) & \rho(k-3) & \dots & \rho(k) \end{bmatrix};$$

dove $\mathbf{P}_{(k)}$ è la matrice di Toeplitz di ordine k , mentre $\mathbf{P}_{(k)}^*$ è la stessa matrice, nella cui ultima colonna è stato sostituito il vettore $(\rho(1), \rho(2), \dots, \rho(k))$ al posto del vettore $(\rho(k-1), \rho(k-2), \dots, 1)$.

Da $\rho(k) = \rho(-k)$ si trova subito $\pi(k) = \pi(-k)$, che porta all'analisi e alla stima di $\pi(k)$ solo per valori di k positivi. Inoltre, applicando la definizione di $\pi(k)$ per i valori di $k = 0, 1, 2, \dots$, si ottiene:

$$\pi(0) = 1 ;$$

$$\pi(1) = \frac{|\rho(1)|}{1} = \rho(1) ;$$

$$\pi(2) = \frac{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & \rho(2) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{vmatrix}} = \frac{\rho(2) - \rho^2(1)}{1 - \rho^2(1)} .$$

Come si vede, nota la funzione di autocorrelazione globale fino al lag k , si può univocamente calcolare la funzione di autocorrelazione parziale fino allo stesso lag k , e viceversa. Infatti si ha anche:

$$\rho(1) = \pi(1) ;$$

$$\rho(2) = \pi(2) + \pi^2(1)[1 - \pi(2)] .$$

È evidente che $\pi(k)$ è funzione di $\rho(k)$ e viceversa, e, quindi, dal punto di vista teorico, $\pi(k)$ non aggiunge nulla alla conoscenza del processo che non sia già teoricamente deducibile da $\rho(k)$.

Per quel che riguarda la stima di $\pi(k)$, questa può essere ricavata direttamente dalle stime precedentemente descritte di $\rho(k)$ attraverso la formula:

$$\hat{\pi}(k) = \frac{|\hat{\mathbf{P}}_{(k)}^*|}{|\hat{\mathbf{P}}_{(k)}|}, \quad k = 1, 2, 3, \dots,$$

dove le matrici $\hat{\mathbf{P}}_{(k)}$ e $\hat{\mathbf{P}}_{(k)}^*$ sono ottenute sostituendo rispettivamente in $\mathbf{P}_{(k)}$ e $\mathbf{P}_{(k)}^*$ gli stimatori $\hat{\rho}(k)$.

Il grafico di $\hat{\pi}(k)$ viene disegnato solo per valori di k positivi ed è usuale stimare l'autocorrelazione parziale per lo stesso numero di lags per i quali è stimata $\rho(k)$, ovvero al più per $N/4$. Tale grafico viene chiamato correlogramma parziale.

Capitolo 2

La classe dei modelli ARIMA

In questo capitolo analizzeremo quella particolare classe di processi stocastici per la quale la ricerca di un modello statistico, a partire dalla conoscenza di una sola serie storica, assume rilevanza operativa e garantisce un'alta probabilità di successo.

Dai dati cercheremo di stimare i parametri di un modello statistico, fra tutti quelli ipotizzabili, in maniera che la serie storica osservata risulti coerente con esso, secondo determinati criteri di adattamento.

In questo capitolo considereremo modelli statistici parametrici che consentono la modellazione di una serie storica attraverso opportune funzioni lineari delle variabili del processo e di White Noise Gaussiani. Analizzeremo dunque i cosiddetti modelli ARMA e le loro opportune trasformazioni.

2.1 – La genesi dei modelli ARMA

Il punto di partenza per arrivare ad una costruzione deduttiva di questi particolari modelli è il teorema di Wold. Esso introduce il processo lineare, una cui approssimazione è il modello “media mobile” (MA, dall'inglese “Moving Average”), a sua volta caratterizzato da un numero finito di autocovarianze se e solo se è “invertibile”.

Ma, se è invertibile, ammette una rappresentazione in funzione delle variabili casuali “passate”, una cui approssimazione è il modello “autoregressivo” (AR, dall'inglese “Auto Regressive”).

Una sintesi delle due classi di modelli conduce ai modelli misti (ARMA). Esaminiamo ora in dettaglio tutti questi passaggi.

2.1.1 – Il Teorema di Wold

Il teorema di Wold afferma: “Ogni processo stocastico stazionario X_t di valor medio μ può essere decomposto in due differenti processi stocastici, stazionari e fra loro mutuamente incorrelati, Z_t e V_t , detti, rispettivamente, componente non deterministica (Z_t) e componente deterministica (V_t), le quali hanno le seguenti rappresentazioni:

$$V_t = \mu + \sum_{j=1}^{\infty} \{ \alpha_j \cos(\lambda_j t) + \beta_j \text{sen}(\lambda_j t) \}$$

$$Z_t = A_t + \psi_1 A_{t-1} + \psi_2 A_{t-2} + \dots, \quad \text{con } \sum \psi_j^2 < +\infty,$$

dove $A_t \sim WN(0, \sigma_a^2)$, mentre $\{\alpha_j\}, \{\beta_j\}$ sono successioni di variabili casuali tali che $E(\alpha_j) = E(\beta_j) = 0$; $Cov(\alpha_j, \beta_j) = 0$, per ogni (i, j) , e $\{\lambda_j\}$ è una successione di numeri reali tali che $0 \leq \lambda_j \leq \pi$, per ogni j ”.

È da notare come il termine “deterministico” nel teorema sia usato in maniera impropria, in quanto in V_t vi sono anche le due variabili casuali $\{\alpha_j\}, \{\beta_j\}$. Questo termine è però lasciato ad indicare che, verificatosi un certo evento, le variabili casuali $\{\alpha_j\}$ e $\{\beta_j\}$ assumeranno i valori numerici $\{\hat{\alpha}_j\}$ e $\{\hat{\beta}_j\}$, per cui V_t sarà un’onda perfettamente periodica, senza alcuna componente stocastica, determinabile con esattezza nel futuro.

Il teorema di Wold afferma quindi che ogni processo stazionario X_t può decomorsi come:

$$X_t = Z_t + V_t,$$

dove Z_t viene anche detto “processo lineare” perchè costituito da una combinazione lineare infinita di processi più semplici (White Noise) e costituisce la parte puramente statistica di X_t , mentre V_t è una combinazione lineare infinita di onde periodiche con ampiezza e fase che sono variabili casuali, ma con frequenza angolare deterministica perchè i λ_j sono quantità fisse.

Essendo poi V_t e Z_t componenti incorrelate, avremo che

$$Var(X_t) = Var(V_t) + Var(Z_t)$$

e, quindi, sarà possibile individuare il peso di V_t nella “spiegazione” del processo X_t .

Se tale peso, misurato da $Var(V_t)/Var(X_t)$, è rilevante sarà necessario estrarre la componente deterministica V_t prima di passare all'analisi residua di $Z_t = X_t - V_t$. Tuttavia la determinazione di V_t a partire da una serie storica (x_1, \dots, x_t) è agevole: si tratta infatti di stimare le costanti $\{a_j\}$ e $\{b_j\}$ nel modello:

$$v_t = \mu + \sum_{j=1}^{\infty} (a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t)) \approx \mu + \sum_{j=1}^m (a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t)),$$

in maniera che sia:

$$\sum_{t=1}^N (x_t - v_t)^2 = \text{minimo}.$$

Questo è un problema di minimi quadrati se sono note le frequenze angolari λ_j .

2.1.2 – Il modello Media Mobile (MA)

È da notare come dal punto di vista teorico sia fondamentale la sommatoria infinita nella definizione di Z_t . Dal punto di vista pratico è evidente che solo un numero finito di termini ψ_1, \dots, ψ_q determina effettivamente le proprietà del processo Z_t , perchè, essendo $\sum \psi_j^2 < \infty$, le quantità ψ_1, ψ_2, \dots dovranno necessariamente tendere a zero e quindi essere trascurabili da un certo punto in poi.

Pertanto non è restrittivo porre:

$$\begin{aligned} \psi_j &= -\theta_j, & \text{per } j = 1, 2, 3, \dots, q \\ \psi_j &= 0, & \text{per } j = q + 1, q + 2, \dots \end{aligned}$$

e considerare quindi il processo stocastico Media Mobile di ordine q , denotato $Z_t \sim MA(q)$, e definito da:

$$Z_t = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}.$$

Ovviamente, essendo un caso particolare del processo lineare, un processo $MA(q)$ è sempre stazionario.

Se la regressione e la conoscenza delle frequenze angolari, che inducono comportamenti rigorosamente periodici, consentono una determinazione esauriente di V_t , l'analisi statistica delle serie storiche si può concentrare sullo studio della componente non deterministica, cioè il processo lineare Z_t .

Il processo Z_t è ovviamente stazionario perchè:

1. $E(Z_t) = 0$,
2. $E(Z_t^2) = Var(Z_t) = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j^2 < +\infty$,
3. $Cov(Z_t, Z_{t+k}) = \gamma(k) = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k}$,

per $k = 0, 1, 2, \dots$ (avendo posto $\psi_0 = 1$),

e, quindi, possiede funzione di autocorrelazione pari a:

$$\rho(k) = \frac{\sum_{j=0}^{\infty} \psi_j \psi_{j+k}}{1 + \sum_{j=1}^{\infty} \psi_j^2}, \quad k = 0, 1, 2, \dots$$

In conclusione un modello $Z_t \sim MA(q)$ è caratterizzato dai $q+1$ parametri $(\sigma_a^2; \theta_1, \dots, \theta_q)$ ovvero dalla varianza di a_t e dai q parametri espliciti del modello.

Un processo $Z_t \sim MA(q)$ è caratterizzato dai $(q+1)$ parametri detti e dalla distribuzione di probabilità delle variabili casuali Z_t .

2.1.3 – Invertibilità nei modelli Media Mobile

Si pone a questo punto un problema: se è noto Z_t si possono calcolare univocamente le autocovarianze $\gamma(0), \gamma(1), \dots$. È vero il contrario? Cioè, note le autocovarianze, è unico il processo Z_t che possiede quelle autocovarianze?

In generale no. Anzi, per ogni successione di autocovarianze $\{\gamma(0), \dots, \gamma(q)\}$, esistono 2^q processi $MA(q)$ che possiedono le stesse fissate autocovarianze. Quindi la funzione di autocovarianza non caratterizza in generale il processo $MA(q)$.

Tuttavia, se si considerano le 2^q q -uple $(\theta_1, \dots, \theta_q)$ differenti, si dimostra che una ed una sola conduce a processi stocastici invertibili.

Considerando solo i processi stazionari e invertibili del tipo $Z_t \sim MA(q)$, per ogni successione $\{\sigma_a^2; \theta_1, \dots, \theta_q\}$ esiste una ed una sola successione $\{\gamma(0), \dots, \gamma(q)\}$ e viceversa.

Pertanto, nella classe dei processi $MA(q)$ invertibili esiste corrispondenza biunivoca fra parametri del modello e funzione di autocovarianza.

2.1.4 – L'operatore backward e il modello Auto Regressivo (AR)

Per trattare la teoria dei processi stazionari ed invertibili è opportuno introdurre ora un operatore lineare B definito attraverso la relazione $BX_t = X_{t-1}$ e, in generale, per ogni $k = 1, 2, \dots$:

$$B^k X_t = B^{k-1}(BX_t) = B^{k-1}X_{t-1} = X_{t-k},$$

con la convenzione $B^0 X_t = X_t$.

L'operatore lineare B è detto backward o di lag e talora è indicato con L .

Tramite l'operatore B , il processo $Z_t \sim MA(q)$ si scrive:

$$\begin{aligned} Z_t &= A_t - \theta_1 B A_t - \theta_2 B^2 A_t - \dots - \theta_q B^q A_t = \\ &= (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) A_t \end{aligned}$$

e, infine:

$$Z_t = \theta(B) A_t,$$

avendo definito con

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

l'operatore media mobile di ordine q .

È allora possibile dimostrare che il processo $Z_t \sim MA(q)$ è invertibile se e solo se tutte le q radici dell'equazione caratteristica, associata al processo $Z_t \sim MA(q)$, definita da:

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q = 0,$$

sono in modulo superiori a 1.

Si può inoltre dimostrare che un processo invertibile $Z_t \sim MA(q)$ può scriversi come

$$Z_t = \pi_1 Z_{t-1} + \pi_2 Z_{t-2} + \dots + \pi_k Z_{t-k} + \dots + A_t,$$

ovvero, utilizzando l'operatore $\pi(B)$ definito da: $\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots$, mediante la formula sintetica

$$\pi(B) Z_t = A_t.$$

Tale struttura viene chiamata “autoregressiva” (AR) perchè paragonabile ad una regressione della variabile Z al tempo t (ovvero Z_t) su se stessa ad un tempo precedente (ovvero Z_{t-1}, \dots, Z_{t-p}).

Naturalmente si può dimostrare che i pesi π_k , $k = 1, 2, \dots$, sono in modulo inferiori ad un prefissato valore positivo, da un certo punto in poi.

Per semplificare l’analisi è dunque ragionevole porre:

$$\begin{aligned} \pi_k &= \phi_k, & \text{per } k = 1, 2, \dots, p, \\ \pi_k &= 0, & \text{per } k = p + 1, p + 2, \dots \end{aligned}$$

In tal modo appare realistico considerare il processo stocastico autoregressivo di ordine p , denotato con $Z_t \sim AR(p)$ e definito dalla relazione:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + A_t$$

Naturalmente, essendo un caso particolare di $\pi(B)Z_t = A_t$, un processo $AR(p)$ è sempre invertibile, ma non è sempre stazionario. Perchè lo sia è necessario imporre dei vincoli sui parametri $\phi_1, \phi_2, \dots, \phi_p$.

Più precisamente, un processo $Z_t \sim AR(p)$ è stazionario se e solo se tutte le p soluzioni dell’equazione caratteristica associata al processo $Z_t \sim AR(p)$ e definita da:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p = 0,$$

sono in modulo superiori a 1.

Utilizzando l’operatore $\phi(B)$ definito sopra, scriviamo la formulazione sintetica di un processo $AR(p)$:

$$\phi(B)Z_t = A_t.$$

Dal punto di vista pratico è interessante notare come dalla formulazione sintetica, attraverso alcuni semplici calcoli, si possa ottenere una equazione omogenea alle differenze finite di ordine p sulla funzione di autocorrelazione $\rho(k)$:

$$\rho(k) = \phi_1 \rho(k-1) + \phi_2 \rho(k-2) + \dots + \phi_p \rho(k-p), \quad k > 0,$$

che, scritta esplicitamente per $k = 1, 2, \dots$, e tenendo conto delle proprietà $\rho(0) = 1$, $\rho(k) = \rho(-k)$, conduce al seguente sistema di equazioni lineari:

oppure si può concepire un processo $W_t \sim MA(q)$ a sua volta connesso ai valori passati di un altro processo $Z_t \sim AR(p)$ secondo le posizioni:

$$\left. \begin{array}{l} W_t = \theta(B) A_t \\ \phi(B) Z_t = W_t \end{array} \right\} \Rightarrow \phi(B) Z_t = \theta(B) A_t .$$

Dal punto di vista della logica delle strutture lineari, si può pensare al processo Z_t come connesso alla sua storia passata (schema AR) con l'interazione di un processo WN A_t a sua volta connesso alla sua serie passata (schema MA). Quindi, mentre lo schema AR è un processo che possiede memoria del passato di Z_t e lo schema MA è un processo che possiede memoria del passato di A_t , uno schema misto del tipo ARMA opera sulla storia del processo Z_t e sulla storia del processo A_t .

Queste motivazioni inducono a considerare il processo stocastico Auto Regressivo di ordine p e Media Mobile di ordine q , indicato con $Z_t \sim ARMA(p, q)$, e definito dalla relazione:

$$Z_t - \phi_1 Z_{t-1} - \phi_2 Z_{t-2} - \dots - \phi_p Z_{t-p} = A_t - \theta_1 A_{t-1} - \dots - \theta_q A_{t-q} ,$$

ovvero, mediante l'introduzione degli operatori AR e MA, dalla relazione:

$$\phi(B) Z_t = \theta(B) A_t .$$

Un processo $Z_t \sim ARMA(p, q)$ equivale, ovviamente, ad un processo $AR(\infty)$ con operatore $\pi(B) = \phi(B)\theta^{-1}(B)$ se è invertibile e ad un processo $MA(\infty)$ con operatore $\psi(B) = \theta(B)\phi^{-1}(B)$ se è stazionario, con pesi π_k e ψ_k ricavabili in maniera ricorsiva.

Si può inoltre dimostrare che un processo $ARMA(p, q)$ è stazionario se tutte le p radici dell'equazione caratteristica associata alla componente AR, cioè $\phi(B) = 0$, sono in modulo superiori a uno, e, inoltre, è invertibile se tutte le q radici dell'equazione caratteristica associata alla componente MA, cioè $\theta(B) = 0$, sono in modulo superiori a uno.

Come si vede, la stazionarietà pone vincoli sui parametri AR (senza legami con quelli MA), mentre l'invertibilità pone vincoli sui parametri MA (senza legami con quelli AR).

2.2 – Serie evolutive e modelli ARIMA

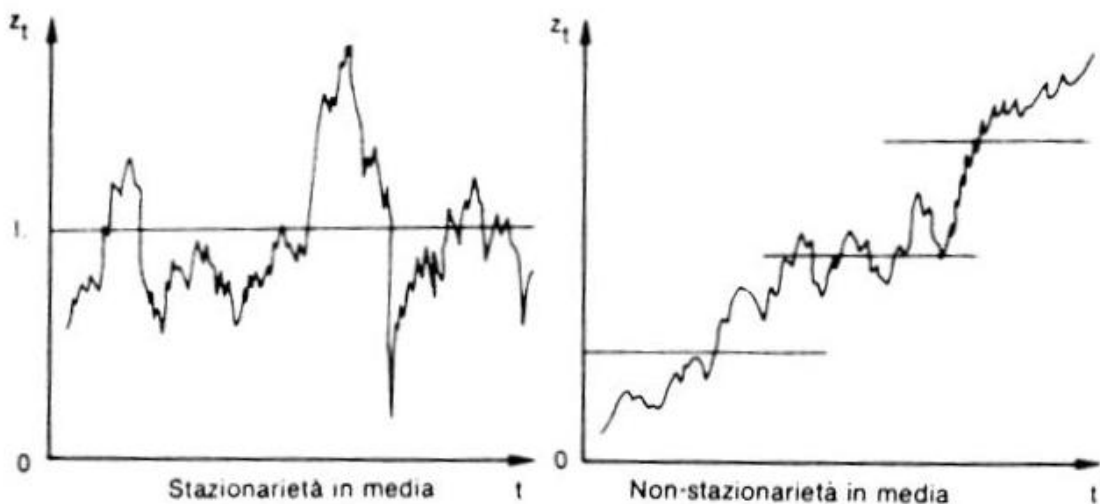
L'introduzione dei modelli ARMA, stazionari ed invertibili, risolve alcuni problemi logici e interpretativi connessi all'analisi delle serie storiche.

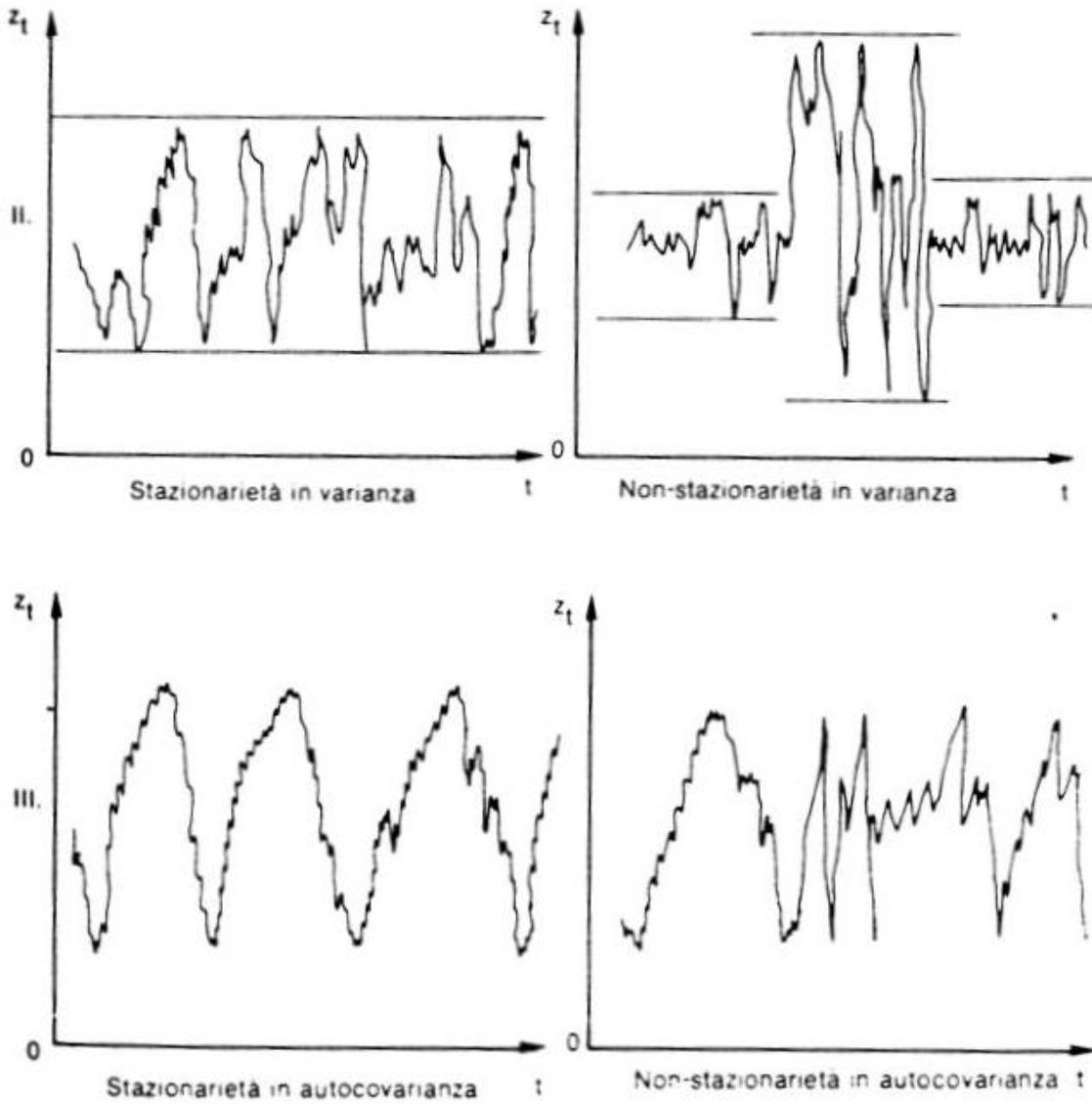
Per tale classe l'individuazione del processo a partire dalla serie può essere ottenuta secondo criteri statisticamente efficienti.

Tuttavia la stazionarietà è un requisito tutt'altro che realistico nelle serie osservate e, specialmente nelle serie economiche, è piuttosto un'eccezione anziché la regola a causa dell'evoluzione della quasi totalità dei fenomeni per i più disparati motivi (incremento demografico e sue conseguenze, inflazione, progresso tecnologico e così via).

Risulta quindi necessario generalizzare i risultati ottenuti sui modelli ARMA per tener conto delle situazioni di non stazionarietà dei dati. Serie storiche non stazionarie (serie evolutive) non possono essere realizzazioni finite di processi stocastici stazionari.

Nelle serie reali la stazionarietà, così come è stata definita in precedenza per i processi stocastici, può essere violata per almeno tre motivi (corrispondenti ai tre vincoli imposti sulla media, sulla varianza e sulla covarianza).





Mentre la non stazionarietà in varianza e covarianza richiedono interventi “ad hoc” (le cosiddette analisi preliminari), la non stazionarietà in media può essere affrontata generalizzando la classe dei processi ARMA.

Supponiamo, per esempio, che un processo Z_t possieda un livello medio μ_t , rappresentabile mediante una retta $\beta_0 + \beta_1 t$. Sia quindi $Z_t = \beta_0 + \beta_1 t + W_t$, essendo W_t una componente stazionaria della cui natura qui non ci occupiamo.

È chiaro che gli incrementi $\mu_t - \mu_{t-1}$ saranno costanti e, ragionando sugli incrementi del processo Z_t , cioè su $Z_t - Z_{t-1}$, si ottiene:

$$\begin{aligned} Z_t - Z_{t-1} &= (\beta_0 + \beta_1 t + W_t) - (\beta_0 + \beta_1(t-1) + W_{t-1}) = \\ &= \beta_1 + W_t - W_{t-1} = \beta_1 + \varepsilon_t, \end{aligned}$$

essendo β_1 un livello medio costante e $\varepsilon_t = W_t - W_{t-1}$ una componente stazionaria come W_t (almeno come media e come varianza).

Ma il processo $Z_t - Z_{t-1}$ può essere anche scritto come:

$$Z_t - Z_{t-1} = Z_t - BZ_t = (1 - B)Z_t ,$$

e, quindi, la differenza prima definita da:

$$\nabla Z_t = Z_t - Z_{t-1} ,$$

non risulta altro che un processo del tipo AR, ma con operatore $\nabla = 1 - B$, la cui radice dell'equazione caratteristica $1 - B = 0$ non è in modulo superiore ad uno (vale esattamente uno).

Se il livello medio μ_t si evolve come una parabola, risulta evidente che per ottenere un livello medio stazionario occorre applicare due volte l'operatore ∇ , ovvero utilizzare la differenza seconda $\nabla^2 = \nabla * \nabla$ definita da:

$$\nabla^d = (1 - B)^d = 1 - dB + \binom{d}{2} B^2 - \dots + (-1)^d B^d .$$

Si nota facilmente che l'operatore AR generalizzato $(1 - B)^2$ possiede due radici esattamente pari a uno: non è quindi stazionario.

Infine se il valor medio del processo si evolve secondo un polinomio di grado d , si può dimostrare che l'applicazione consecutiva d volte dell'operatore ∇ , cioè ∇^d , genera un processo stazionario in media.

In tal caso l'operatore AR generalizzato definito da:

$$\nabla^d = (1 - B)^d = 1 - dB + \binom{d}{2} B^2 - \dots + (-1)^d B^d ,$$

possiede tutte le d radici esattamente pari a uno.

Pertanto dopo avere applicato d volte l'operatore ∇ al processo non stazionario Z_t , si otterrà un processo W_t che sarà stazionario e che può ammettere, per quanto detto in precedenza, una rappresentazione ARMA.

Formalmente, sia Z_t un processo non stazionario tale che $\nabla^d Z_t = W_t$ sia stazionario e per il quale valga una rappresentazione ARMA stazionaria e invertibile; allora si avrà:

$$\left. \begin{array}{l} \nabla^d Z_t = W_t \\ \phi(B)W_t = \theta(B)A_t \end{array} \right\} \Rightarrow \nabla^d \phi(B)Z_t = \theta(B)A_t .$$

Tali risultati ci inducono a considerare quindi il processo stocastico Auto Regressivo di ordine p , Integrato d volte e Media Mobile di ordine q , denotato con $Z_t \sim ARIMA(p, d, q)$ e definito dalla relazione:

$$\nabla^d \{Z_t - \phi_1 Z_{t-1} - \dots - \phi_p Z_{t-p}\} = A_t - \theta_1 A_{t-1} - \dots - \theta_q A_{t-q},$$

ovvero, equivalentemente, dalla relazione sul processo $W_t = \nabla^d Z_t$:

$$W_t - \phi_1 W_{t-1} - \dots - \phi_p W_{t-p} = A_t - \theta_1 A_{t-1} - \dots - \theta_q A_{t-q}.$$

La motivazione che Box e Jenkins forniscono al termine “integrato” per tali processi è che per riottenere Z_t da W_t occorre sommare (e quindi integrare) rispetto a W_t .

Quindi in generale:

$$Z_t \sim ARIMA(p, d, q) \Leftrightarrow \nabla^d Z_t = W_t \sim ARMA(p, q)$$

e, tramite gli operatori lineari, sarà possibile scrivere:

$$\nabla^d \phi(B)Z_t = \theta(B)A_t \Leftrightarrow \phi(B)W_t = \theta(B)A_t,$$

ovvero, equivalentemente:

$$\phi(B)Z_t = \theta(B)A_t,$$

avendo definito l'operatore AR generalizzato di ordine $p + d$:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_{p+d} B^{p+d} = \phi(B)(1 - B)^d.$$

Si noti che, tramite l'operatore $\phi(B)$, il processo ARIMA è ancora un processo con formulazione ARMA, con operatore AR non stazionario, ma con d delle $p + d$ radici dell'equazione caratteristica $\phi(B) = 0$ esattamente pari a uno.

E' da notare che noi imponiamo che le radici di $\phi(B)$ siano in modulo pari a uno o in modulo superiori a uno. Si escludono, quindi, i processi con radici in modulo inferiori a uno che danno luogo a realizzazioni “esplosive” con $\mu_t \rightarrow \infty$.

Capitolo 3

La costruzione di un modello ARIMA

3.1 – Identificabilità dei modelli ARMA

Per identificabilità si intende la possibilità di stimare in modo univoco la distribuzione del processo stocastico a partire dalla serie storica osservata, nell'ambito di una prefissata classe di distribuzioni.

Cercheremo dunque di mostrare che è possibile esaurire il ciclo “serie \rightarrow modello \rightarrow processo” secondo una procedura statisticamente efficiente.

Nota la serie storica $\{x_t, t = 1, \dots, N\}$ si stima, come primo passo, la funzione di autocorrelazione (globale e parziale) mediante gli stimatori adeguati.

La conoscenza di $\rho(k)$ e della sua trasformata $\pi(k)$ non è generalmente sufficiente né per il modello, né per il processo stesso. Limitandosi, però, a modelli ARMA stazionari e invertibili, si crea una corrispondenza biunivoca fra modello e funzione di autocorrelazione.

D'altra parte il modello ARMA e la funzione $\rho(k)$ non esauriscono la conoscenza del processo X_t , di cui quel modello è una rappresentazione, a meno di non specificare una distribuzione di probabilità per le variabili casuali componenti il processo X_t .

Si è visto che considerando processi Gaussiani stazionari (di valor medio nullo) la conoscenza della funzione di autocovarianza equivale alla conoscenza del processo.

Pertanto la funzione di autocorrelazione (proporzionale tramite $\gamma(0)$ alla funzione di autocovarianza) è caratterizzante per processi Gaussiani.

3.2 – Costruzione iterativa di un modello ARIMA

La procedura che vedremo è quella proposta da Box e Jenkins, con la quale si cerca di ottenere un'analisi di una qualsiasi serie storica, attraverso varie fasi e vari strumenti statistici.

Il primo passo per la costruzione di un modello ARIMA è rappresentato dalle analisi preliminari. Per mezzo di esse si cerca di operare sulla serie originaria per ricondurla alla realizzazione di un processo stazionario e Gaussiano.

Nella fase seguente, tramite la funzione di autocorrelazione (globale e parziale), si cerca di pervenire all'identificazione del modello ARIMA, cioè alla determinazione degli ordini (p, d, q) che, rispettivamente, caratterizzano gli operatori $\phi(B), \nabla, \theta(B)$ del modello.

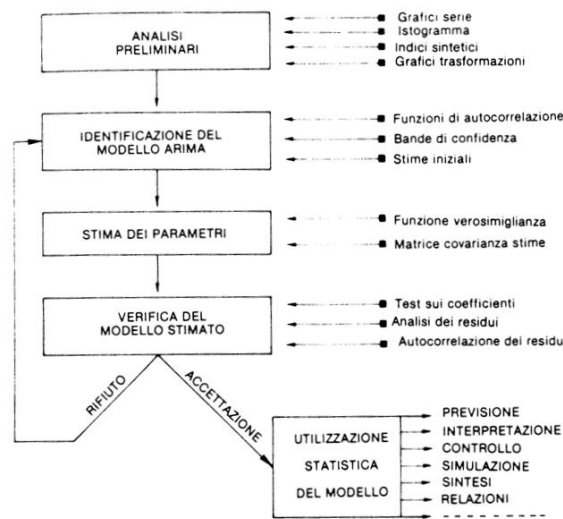
La fase successiva è la stima dei parametri, che determina, in maniera tipicamente automatica, sotto l'ipotesi di normalità del processo $A_t \sim WN$, la stima e l'errore standard delle stime dei parametri del modello ARIMA precedentemente identificato.

Una verifica del modello stimato avviene tramite l'analisi statistica dei residui stimati mediante quel modello.

L'accettazione dell'ipotesi $A_t \sim WN$ equivale all'accettazione del modello, mentre il rifiuto implica il rigetto del modello stimato e la ricerca di un modello più adeguato.

Proprio in caso di rifiuto, la ripetizione del ciclo "identificazione - stima - test" viene effettuata sulla base dei motivi che hanno condotto al rifiuto di quella ipotesi principale.

Quindi il rifiuto del modello dà indicazioni su come costruire un modello alternativo.



Non esistono però motivazioni teoriche per cui il ciclo iterativo debba concludersi dopo un numero finito di modelli alternativi rifiutati.

3.2.1 – Analisi Preliminari

Nei capitoli precedenti abbiamo visto che per serie storiche che possono essere pensate come realizzazioni finite di processi Gaussiani stazionari e invertibili, la procedura fondata sulla funzione di autocorrelazione per modelli ARMA costituisce uno strumento statistico efficiente per risalire dalla serie storica al processo stocastico.

Purtroppo solo raramente si è in presenza di serie stazionarie, per cui ci si deve porre il problema di verificare fino a che punto sia possibile trasformare la serie originaria per ricondurla alle realizzazioni tipiche di processi stazionari e Gaussiani.

Ciò implica una serie di verifiche che chiamiamo analisi preliminari, tendenti ad accertare almeno tre condizioni:

1. distribuzione normale del WN A_t ,
2. assenza e eliminazione di valori anomali (outliers),
3. stazionarietà in media, varianza e covarianza .

Per quel che riguarda il primo punto è sufficiente verificare la normalità del processo Z_t (che, essendo stazionario, è combinazione lineare di A_t) mediante test di accostamento alla distribuzione normale condotti sulla serie storica degli scarti $z_t = x_t - \bar{x}$, $t = 1, \dots, N$. In pratica non si costruisce un'analisi molto raffinata, ma ci si limita a costruire l'istogramma di frequenza per z_t , accertandosi che sia almeno simmetrico e unimodale, poiché, per tali distribuzioni di frequenza, la procedura di Box e Jenkins viene ritenuta statisticamente robusta.

L'analisi degli outliers avviene principalmente perchè un dato anomalo aumenta, anche considerevolmente, la varianza e l'autocovarianza a tutti i lag e, dunque, può portare a problemi durante l'identificazione del modello ARMA.

La sostituzione avviene tramite metodologie specifiche, come l'Intervention Analysis introdotta nel 1975 da Box e Tiao, che non solo non sostituisce arbitrariamente il dato anomalo, ma consente una possibile valutazione dei suoi effetti sulla dinamica della serie.

Il problema principale resta quindi quello di costruire realizzazioni stazionarie a partire da serie storiche che generalmente non lo sono.

Per la non stazionarietà in media (trend) si può operare sulla serie originaria x_t in vari modi. Tra questi il più utilizzato consiste nel costruire la serie z_t delle differenze di ordine $d = 1, 2, 3, \dots$ scegliendo la differenza d -esima per la quale le realizzazioni di $\nabla^d z_t = w_t$ risultano ad un'analisi grafica, abbastanza vicine alla stazionarietà rispetto al livello medio.

Tale metodo risulta piuttosto semplice e meccanico, ma ha lo svantaggio di far perdere informazioni al crescere di d e di indurre autocorrelazioni negative nelle componenti MA.

Per la non stazionarietà in varianza (eteroschedasticità) si opera generalmente seguendo due possibili metodi.

Per gli studi econometrici si preferisce trasformare dati sempre positivi (o traslati perchè risultino sempre positivi) mediante il logaritmo, dato che tale funzione diminuisce il divario fra minimi e massimi di una successione. In presenza di trend in media (eliminabile con l'operatore ∇) e di trend in varianza (riducibile con la funzione logaritmo) è piuttosto comune trasformare la serie x_t nella serie $w_t = \nabla \log(x_t)$ perchè il processo $W_t = \nabla \log(X_t)$ è, approssimativamente, il tasso di variazione relativo di X_t .

In generale, invece, seguendo la logica propria dell'analisi delle serie storiche, la ricerca di una trasformazione ottimale per i dati viene impostata in modo che sia la serie storica a fornire indicazioni sulla trasformazione ottimale.

Nel 1964 Box e Cox hanno introdotto una trasformazione che da loro prende anche il nome. La trasformazione di Box-Cox per il processo Z_t con parametro λ è il processo $Z_t(\lambda)$ definito da:

$$Z_t(\lambda) = \begin{cases} \frac{Z_t^\lambda - 1}{\lambda}, & \text{per } \lambda \neq 0, \\ \log(Z_t), & \text{per } \lambda = 0. \end{cases}$$

Data la serie storica (z_1, \dots, z_t) si cerca il valore λ^* di λ tale che $z_t(\lambda^*)$ sia la più omoschedastica possibile.

In generale, ipotizzato un modello $ARMA(p, q)$, per la serie occorrerebbe cercare stime congiunte dei parametri AR, dei parametri MA, di σ_a^2 e di λ .

Si nota che la trasformazione di Box-Cox è continua nell'origine in quanto $\lim_{\lambda \rightarrow 0^+} Z_t(\lambda) = \log Z_t$ e include praticamente tutte le trasformazioni più utilizzate in letteratura (logaritmica per $\lambda = 0$, inversa per $\lambda = -1$, radice quadrata per $\lambda = 0.5$, quadrato per $\lambda = 2$ e così via) e altre difficilmente individuabili dall'analisi grafica. È da notare che se $\lambda \approx 1$, tale analisi suggerirebbe di non trasformare la serie.

I tempi e i costi di una elaborazione congiunta sui parametri ARMA e sul parametro λ sono notevoli, per cui esistono vari metodi approssimati proposti per ottenere stime più veloci.

Infine, se non è possibile assumere la stazionarietà in covarianza, si può procedere, possedendo molti dati, alla suddivisione della serie originaria in sottoserie a covarianza omogenea.

3.2.2 – L'identificazione del modello

Avendo risolto la questione di poter modellare serie stazionarie, supporremo d'ora in poi che z_t sia una realizzazione finita, per $t = 1, \dots, N$, stazionaria e di valor medio zero.

Di questa calcoliamo le autocovarianze $\gamma(k)$ e la funzione di autocorrelazione $\rho(k)$ tramite gli stimatori precedentemente descritti:

$$\hat{\gamma}(k) = \frac{1}{N} \sum_{t=1}^{N-k} z_t z_{t+k} \quad \text{e} \quad \hat{\rho}(k) = \hat{\gamma}(k) / \hat{\gamma}(0) \quad \text{per } k = 0, 1, 2, \dots$$

Dallo studio della funzione di autocorrelazione globale $\hat{\rho}(k)$ e della funzione di autocorrelazione parziale $\hat{\pi}(k)$, ovvero dal confronto dell'andamento osservato rispetto all'andamento (teorico) deducibile per processi $AR(p)$, processi $MA(q)$ e processi $ARMA(p, q)$ occorre decidere l'ordine p e q delle componenti AR e MA.

È dunque di fondamentale importanza sapere quali sono i comportamenti tipici e teorici delle due funzioni per i processi $ARMA(p, q)$ al variare di p e q .

Processo stocastico stazionario e invertibile	Funzioni di autocorrelazione	
	Globale $\rho(k)$	Parziale $\pi(k)$
$AR(p)$	Non si annulla mai, ma decade verso lo zero	È nulla per $k > p$
$MA(q)$	È nulla per $k > q$	Non si annulla mai, ma decade verso lo zero
$ARMA(p, q)$	Non si annulla mai, ma decade verso lo zero per i lags $k > p - q$	Non si annulla mai, ma decade verso lo zero per i lags $k > q - p$

La conoscenza dei comportamenti tipici per i differenti processi, l'esperienza e alcune osservazioni teoriche (che qui non riportiamo) ci guidano dunque verso la scelta degli ordini p e q che caratterizzano il nostro modello ARMA.

3.2.3 – La stima dei parametri

Identificati gli ordini (p, q) del modello ARMA definito da:

$$W_t = \phi_1 W_{t-1} + \dots + \phi_p W_{t-p} + A_t - \theta_1 A_{t-1} - \dots - \theta_q A_{t-q},$$

dove $A_t \sim WN(0, \sigma_a^2)$, occorre procedere ad una stima efficiente dei $(p + q + 1)$ parametri che lo caratterizzano, cioè dal vettore

$$\beta' = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma_a^2).$$

Ipotizzando una distribuzione Gaussiana per A_t e, quindi, per W_t , il metodo della massima verosimiglianza fornisce stimatori con proprietà statisticamente ben definite.

Dal punto di vista pratico è quasi universalmente utilizzato l'algoritmo di Marquardt per l'analisi delle serie storiche, che costituisce una sorta di algoritmo medio fra lo "steepest descent" e il metodo di Newton-Raphson. Inoltre è di grande aiuto, per rendere più veloce il calcolo, fornire valori iniziali abbastanza prossimi al vettore ottimale $\hat{\beta}'$ cercato.

3.2.4 – La verifica del modello

Ottenute le stime dei parametri di un modello $ARMA(p, q)$, si procede ad un'analisi sui residui in modo che si possa verificare la correttezza del modello stimato. In altri termini, si cerca di controllare se i residui risultino effettivamente casuali e normali, secondo l'ipotesi assunta per il WN A_t .

Tale verifica segue tre passi principali:

- a) *analisi grafica dei residui*: serve per individuare andamenti e valori anomali da correggere. Tale analisi, in caso di correttezza del modello, dovrebbe concludere che i residui sono distribuiti nel grafico in maniera omogenea.
- b) *analisi della distribuzione dei residui*: si costruisce, generalmente, l'istogramma dei residui per verificarne la normalità o, almeno, la simmetria e unimodalità. Si possono inoltre applicare i test usuali sulla forma della distribuzione e sulla bontà di adattamento mediante il test del χ^2 .
- c) *stima dell'autocorrelazione dei residui*: poiché l'ipotesi $A_t \sim WN$ imporrebbe una funzione di autocorrelazione $\rho_A(k) = 0$ per ogni $k \neq 0$, si calcola la funzione di autocorrelazione $\hat{\rho}_a(k)$ dei residui stimati \hat{a}_t , che presenta varianza approssimata $1/N$ e si costruiscono quindi bande di confidenza attorno al valore zero. Ogni valore al di fuori della banda ovvero ogni andamento caratteristico di componenti AR o MA è ritenuto indicativo di inadeguatezza del modello.

Quest'ultimo passo è il più rilevante nell'ambito della proposta di Box e Jenkins, in quanto indica la strada da seguire per la ricerca di modelli alternativi a quello analizzato. Supponiamo, per esempio, che il modello $ARMA(p, q)$, stimato per la serie storica z_t , sia $\hat{\phi}_p(B)z_t = \hat{\theta}_q(B)\hat{a}_t$, ma che dall'analisi della funzione di autocorrelazione dei residui \hat{a}_t si debba dedurre che $A_t \sim ARMA(r, s)$, cioè $\hat{\phi}'_p(B)A_t = \hat{\theta}'_q(B)E_t$, dove $E_t \sim WN$, e supponiamo che ciò sia confermato nella fase di verifica.

Il modello complessivo da identificare, stimare e verificare nel successivo ciclo iterativo sarà quindi:

$$\hat{\phi}_p(B)\hat{\phi}_r'(B)Z_t = \hat{\theta}_q(B)\hat{\theta}_s'(B)E_t,$$

e, quindi, utilizzeremo la classe dei modelli $ARMA(p+r, q+s)$.

Un ulteriore aiuto nella scelta del modello è dato dai test globali sulle autocorrelazioni dei residui. Con essi si cerca di costruire una regione critica per l'ipotesi $H_0 : A_t \sim WN$, contro ogni alternativa differente. Uno dei migliori test è il test di Ljung e Box, basato sulla statistica:

$$Q = N(N+2) \sum_{k=1}^m (N-k)^{-1} [\rho_a(k)]^2,$$

che ha distribuzione asintotica $\chi_{[m-(p+q+1)]}^2$.

Capitolo 4

Applicazioni a dati economici

4.1 – Descrizione del problema

Cercheremo ora di analizzare due serie storiche di ambito economico per poter sviluppare un modello ARIMA che espliciti il comportamento dei dati nel tempo. Tale modello può essere utilizzato sia per descrivere l'andamento dei dati nel tempo trascorso, sia per cercare di prevederne il comportamento futuro.

L'analisi statistica è stata realizzata tramite il software SAS ed in particolare mediante la “proc arima”.

4.2 – Descrizione della procedura ARIMA

La *proc arima* è una procedura interattiva che riprende le tre fasi di modellazione dei processi ARIMA di Box e Jenkins:

- la fase di *identificazione*, durante la quale si analizzano le funzioni di autocorrelazione per trarre qualche indicazione sui parametri del modello, limitando così il numero di modelli plausibili, senza però necessariamente identificare il modello migliore;
- la fase di *stima e verifica diagnostica*, durante la quale si stimano i parametri del modello tramite algoritmi iterativi, si giudica l'adeguatezza del modello tramite opportuni test sull'autocorrelazione dei residui e sulla significatività dei parametri, e tramite criteri di informazione;

- la fase di *previsione*, durante la quale si utilizza il modello che si è accettato per produrre le previsioni e i relativi intervalli di confidenza dei valori futuri della serie.

Gli statement relativi a tali fasi sono, rispettivamente: *identify*, *estimate* e *forecast*.

4.2.1 – Lo statement identify

Durante la fase di identificazione, SAS riporta i valori delle funzioni di autocorrelazione, autocorrelazione inversa (ossia la funzione di autocorrelazione del processo duale) e autocorrelazione parziale campionarie e le relative deviazioni standard fino ad un intervallo specificato con l'opzione *nlag*. Se non specificato di default la procedura riporta $\min(n/4, 24)$ valori, dove n è il numero di dati, in quanto per $k > n/4$ le stime delle funzioni di autocorrelazione diventano poco significative. Infine SAS riporta anche un test globale basato sull'autocorrelazione dei residui sotto l'ipotesi principale $H_0 : A_t \sim WN$.

Tre opzioni interessanti associate a questo statement sono *esacf*, *minic* e *scan*. Con esse si cerca di trovare l'ordine degli operatori autoregressivi e a media mobile tramite la funzione di autocorrelazione campionaria estesa (Extended Sample AutoCorrelation Function), la minimizzazione del criterio di informazione di Akaike (MINimization of the Information Criteria) o il calcolo delle stime della correlazione (estimates of the Squared CANonical correlations).

4.2.2 – Lo statement estimate

Durante la fase di stima e verifica diagnostica SAS visualizza in output:

- le stime dei parametri, le relative deviazioni standard e i valori *t*-ratio, che permettono di valutare la significatività dei parametri;
- la stima della varianza e della deviazione standard del processo White Noise e i criteri di informazione AIC (criterio di Akaike) e SBC (criterio di Schwartz bayesiano, chiamato anche BIC), fondamentali per giudicare la bontà di adattamento del modello;

- le correlazioni delle stime dei parametri, che permettono di individuare i parametri ridondanti;
- il controllo sull'autocorrelazione dei residui, ovvero le autocorrelazioni dei residui con le relative statistiche e p-value;
- il modello stimato, ovvero le componenti autoregressive e a media mobile.

I metodi di stima utilizzano le stime preliminari ottenute tramite le funzioni di autocorrelazione campionarie; il SAS fornisce tre possibili metodi:

- il *metodo CLS* (minimi quadrati condizionali) che minimizza la somma dei quadrati dei residui assumendo che i valori passati della serie siano uguali alla media campionaria;
- il *metodo ULS* (minimi quadrati non condizionali) che minimizza la somma dei quadrati dei residui;
- il *metodo ML* (massima verosimiglianza) che massimizza la verosimiglianza.

4.2.3 – Lo statement forecast

In fase di previsione SAS visualizza:

- il modello stimato;
- le previsioni, le relative deviazioni standard e gli intervalli di confidenza che, di default, hanno livello $1 - \alpha = 0.95$.

Le previsioni e i limiti di confidenza, come già detto, sono legate al metodo di stima. Se però il modello è adeguato e il numero di osservazioni è sufficientemente grande, le previsioni ottenute con i tre diversi metodi non presentano in genere differenze apprezzabili.

In fase di verifica diagnostica si studiano principalmente l'autocorrelazione dei residui e la significatività delle stime dei parametri e le correlazioni fra gli stessi.

I criteri di informazione vanno confrontati con quelli di altri modelli, stimati con lo stesso metodo.

4.3 – Dati analizzati

Gli esempi che proporremo sono due: nel primo utilizzeremo i dati riguardanti la serie storica del prezzo del petrolio al barile in dollari USA (più precisamente il “Spot Oil Price: West Texas Intermediate”) dal gennaio del 1946 all’ottobre del 2004 reperiti su <http://www.hussmanfunds.com/html/datapage.htm>; nel secondo esempio analizzeremo la produzione mensile di elettricità in Australia dal gennaio del 1956 all’agosto del 1995 (dati reperiti su <http://www-personal.buseco.monash.edu.au/~hyndman/TSDL>).

In entrambi i datasets sono presenti due sole variabili:

- **value**, che rappresenta il valore della variabile da analizzare (dollari al barile o m.KWH di elettricità prodotta),
- **date**, che rappresenta il mese e l’anno di riferimento del valore.

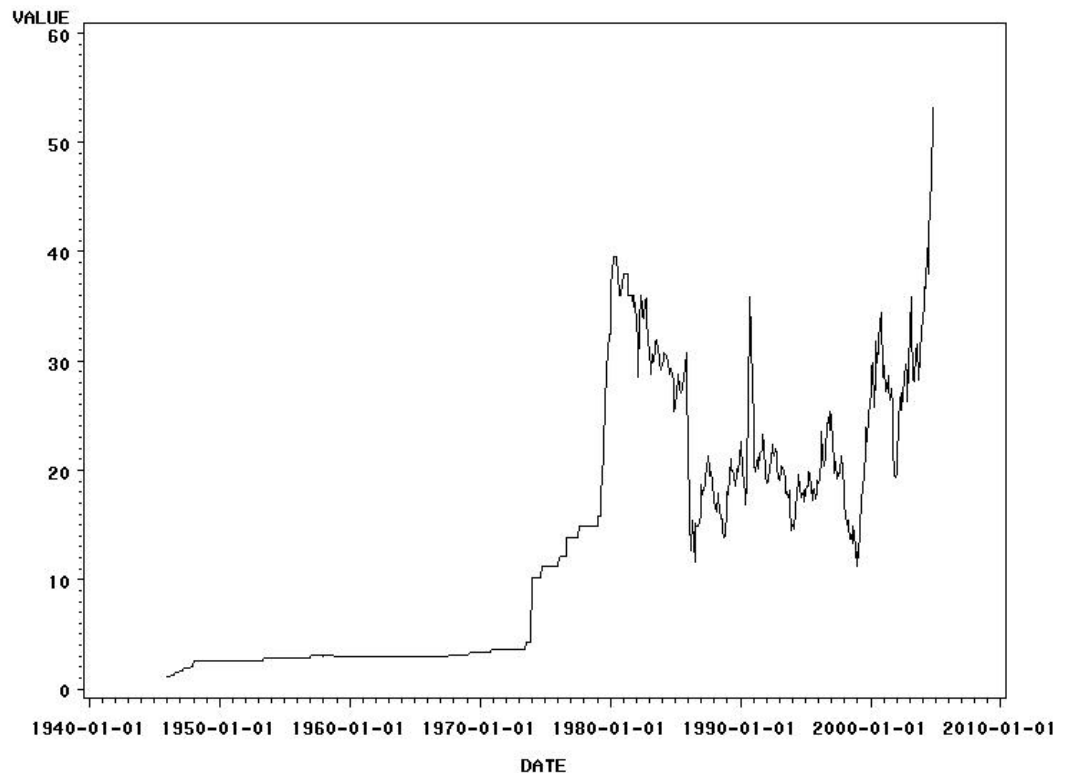
4.4 – Analisi del prezzo del petrolio al barile

Cominciamo innanzi tutto col mostrare i dati in forma grafica.

```
libname dati 'D:\Home\Tesi\libreria';
proc import out=dati.oil
    datafile="D:\Home\Tesi\Dati\oil.txt"
    DBMS=DLM replace;
    delimiter='20'x;
    getnames=YES;
    datarow=2;
run;

proc gplot data=dati.oil;
    plot value*date=1;
    symbol1 line=1 c=black i=join;
    title 'Prezzo del petrolio al barile';
run;
quit;
```

Prezzo del petrolio al barile



Notiamo immediatamente che l'andamento dei dati può sembrare piuttosto singolare: per venti anni il dato è quasi costante, ma improvvisamente ha un picco in salita (gennaio 1974), seguito qualche anno più tardi da un secondo picco (da maggio del 1979 ad aprile del 1980). A questo punto il prezzo tende a scendere e notiamo addirittura un crollo del prezzo (gennaio-marzo del 1986) che riporta il costo ai livelli di otto anni prima. Seguono tredici anni di grande instabilità (con un picco molto evidente nell'autunno del 1990), alla fine dei quali inizia una salita che porta ad un valore massimo (53.13) proprio all'ultimo mese rilevato.

Questo andamento provoca dunque una evidente non stazionarietà in media e in varianza. Per cercare di smorzare questi effetti negativi, trasformiamo la nostra serie x_t nella serie $\nabla \text{Log}(x_t)$, così come è usuale per gli studi econometrici. Appliciamo dunque il logaritmo ai nostri dati assieme all'operatore Delta di ordine uno e utilizziamo la proc arima di SAS per la ricerca di un modello adeguato.

Per mezzo delle tre opzioni "scan", "esacf" e "minic" cerchiamo l'ordine degli operatori Autoregressivi e a Media Mobile:

```
proc arima data=dati.logoil;
    identify var=value(1) esacf p=(0:10) q=(0:10) nlag=48;
run;
    identify var=value(1) minic p=(0:10) q=(0:10) nlag=48;
run;
    identify var=value(1) scan p=(0:10) q=(0:10) nlag=48;
run;
quit;
```

Nell'output di SAS possiamo leggere i principali candidati:

ARMA(p+d,q) Tentative Order Selection Tests		<i>Minimum Table Value:</i> $BIC(1,0) = -5.57241$	ARMA(p+d,q) Tentative Order Selection Tests	
ESACF			SCAN	
p+d	q		p+d	q
1	1	1	1	
9	8	9	0	
0	9	0	9	

Esaminiamo quindi con lo statement estimate le differenti possibilità:

```
proc arima data=dati.logoil;
    identify var=value(1) esacf p=(0:10) q=(0:10) nlag=48 noprint;
run;
    estimate p=1 q=1;
run;
    estimate p=9 q=8;
run;
    estimate p=0 q=9;
run;
    estimate p=1 q=0;
run;
    estimate p=9 q=0;
run;
run;
quit;
```


Per confrontare i modelli utilizziamo principalmente i criteri di bontà AIC e SBC. Durante il confronto è molto utile analizzare anche la stima e la significatività dei singoli parametri, sia per mezzo di test, sia tramite la correlazione tra parametri.

Conditional Least Squares Estimation					
Parameter	Stima	Standard Error	Valore t	Approx Pr > t	Lag
MU	0.0054496	0.0027818	1.96	0.0505	0
MA1,1	-0.22477	0.19000	-1.18	0.2372	1
AR1,1	-0.03077	0.19483	-0.16	0.8746	1

Correlations of Parameter Estimates			
Parameter	MU	MA1,1	AR1,1
MU	1.000	0.002	0.003
MA1,1	0.002	1.000	0.981
AR1,1	0.003	0.981	1.000

Possiamo per esempio notare dalle due tabelle precedenti, come per il modello $ARMA(1,1)$, entrambi i parametri siano singolarmente eliminabili, sia che si consideri il test sulla nullità dei parametri, sia considerando l'alta correlazione. Il nostro interesse è infatti ricercare un modello che risulti buono e sia il più semplice possibile.

Viste le considerazioni sui modelli iniziali, testiamo altri modelli, e scegliamo uno di essi, ossia il modello $ARMA(0,1)$, che sembra essere il migliore.

Innanzitutto notiamo che il parametro $MA(1,1)$ non può essere considerato non significativo:

Conditional Least Squares Estimation					
Parameter	Stima	Standard Error	Valore t	Approx Pr > t	Lag
MU	0.0054504	0.0027951	1.95	0.0516	0
MA1,1	-0.19488	0.03714	-5.25	<.0001	1

Nell'output di SAS è interessante notare anche il controllo sulla autocorrelazione dei residui:

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	3.84	5	0.5721	-0.001	-0.002	0.001	-0.048	0.028	-0.048
12	10.55	11	0.4816	-0.002	0.025	0.083	0.009	0.040	-0.015
18	26.15	17	0.0718	-0.074	0.016	-0.100	-0.073	0.021	0.010
24	30.39	23	0.1384	0.013	0.057	0.030	0.009	-0.031	-0.020
30	37.10	29	0.1438	-0.026	-0.067	-0.020	0.051	0.004	0.030
36	42.50	35	0.1792	0.029	0.026	-0.025	-0.066	-0.006	0.027
42	46.41	41	0.2592	0.049	0.002	-0.028	-0.006	0.034	0.028
48	48.69	47	0.4050	-0.004	0.005	0.023	-0.037	0.026	0.020

Per ogni lag considerato non possiamo rifiutare l'ipotesi principale che i residui siano incorrelati. Questo ci porta a dire che il nostro modello, che ha forma esplicita:

$$\nabla \text{Log}(X_t) = (1 + \theta_1 B)Z_t,$$

con $\hat{\theta}_1 = 0,19488$ e $Z_t \sim WN(0, \sigma^2)$, è adeguato.

Creiamo dunque il nostro modello con lo statement forecast e vediamo come si comporta nel descrivere i nostri dati e nel prevedere l'andamento futuro della variabile "value".

```
proc arima data=dati.logoil;
    identify var=value(1) esacf p=(0:10) q=(0:10) nlag=48 noprint;
    estimate p=0 q=1;
    forecast alpha=0.05 lead=48 out=dati.outlogoil
        id=date interval=month;
```

```
run;
```

```
quit;
```

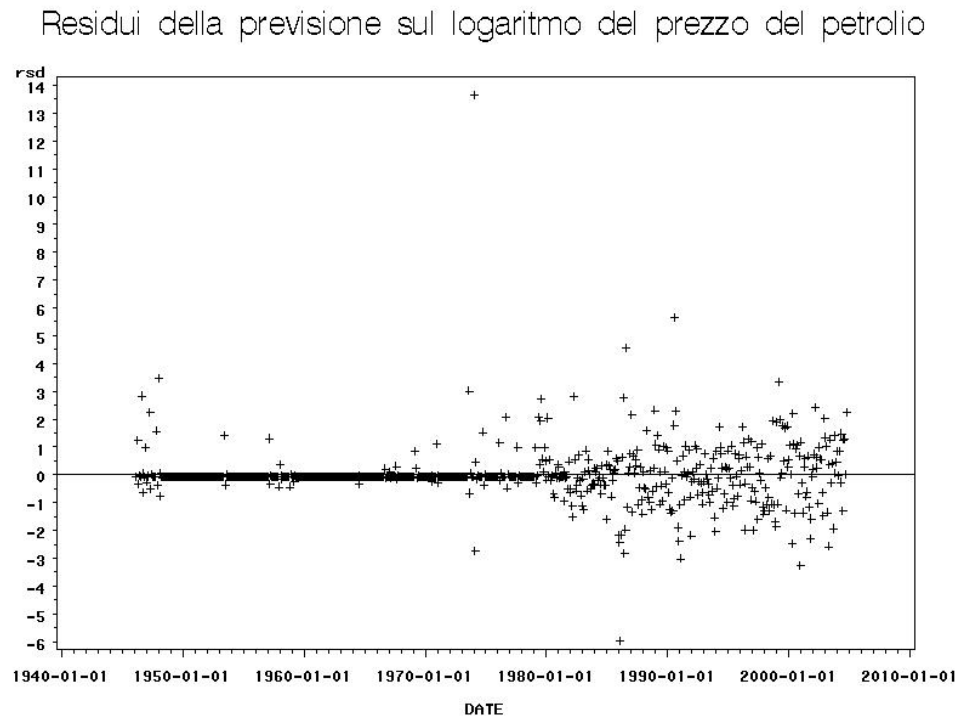
```
data temp;
```

```
set dati.outlogoil;
```

```
rsd=residual/std;
```

```
run;
```

```
proc gplot data=temp;
  plot (rsd)*date=1 / vref=0;
  symbol1 c=black v=plus i=none;
  title 'Residui della previsione sul logaritmo del prezzo del petrolio';
run;
quit;
```

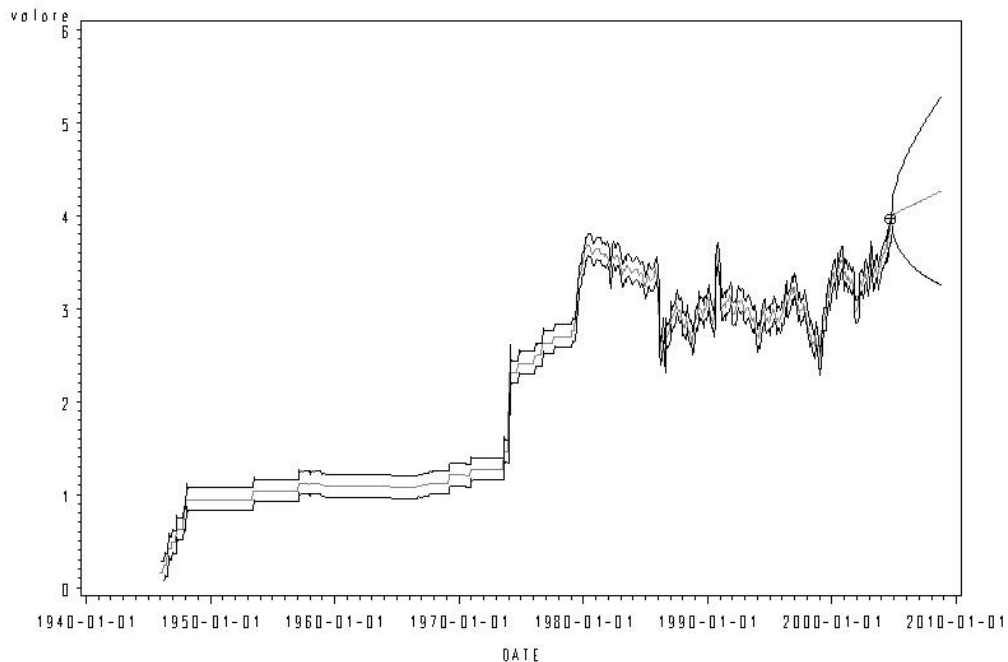


Il grafico dei residui non mostra casualità e presenta valori ben al di sopra del consueto limite $(-3,+3)$, soprattutto nei momenti di grande salita o grande discesa della variabile in oggetto: i due valori più estremi corrispondono infatti all'anno 1974 (*residuo* = 13,65) e all'anno 1986 (*residuo* = -5,99). Considerando però l'estrema variabilità dei nostri dati, soprattutto dal 1979 in poi, ci possiamo ritenere soddisfatti del modello.

Osserviamo ora le previsioni che esso ci fornisce entro i limiti di confidenza del 95% :

```
proc gplot data=tempbis;
  plot valore*date=1 zz*date=2 l95*date=3 u95*date=4 / overlay;
  symbol1 line=1 c=grey i=join;
  symbol2 c=black v='+';
  symbol3 line=1 c=black i=join;
  symbol4 line=1 c=black i=join;
  title 'Intervalli di confidenza sul prezzo del petrolio';
run;
quit;
```

Intervalli di confidenza sul prezzo del petrolio



Per quel che riguarda i dati a nostra disposizione, il modello sembra dunque comportarsi piuttosto bene nei periodi di relativa stabilità, ma ha grossi limiti nei periodi con alta variabilità. Per quel che riguarda la previsione (per un periodo di 48 mesi, ovvero quattro anni) notiamo come il modello ci dica che il prezzo continuerà a salire, senza però fornirci una indicazione concreta: nel momento in cui si cerca di prevedere il futuro si nota una “esplosione” dell’intervallo di confidenza.

Si otterrebbe una migliore stima modellando la serie storica “a tratti”, ossia tagliando i dati in corrispondenza dei picchi osservati, per ottenere migliori risultati sia come analisi che come previsione.

Questo modo di procedere sarebbe però praticamente inutile perchè ogni modello sarebbe “a sé stante”, perdendo significato soprattutto per la previsione: nel momento in cui si verifica un nuovo picco il modello creato sui dati precedenti risulterebbe infatti inapplicabile.

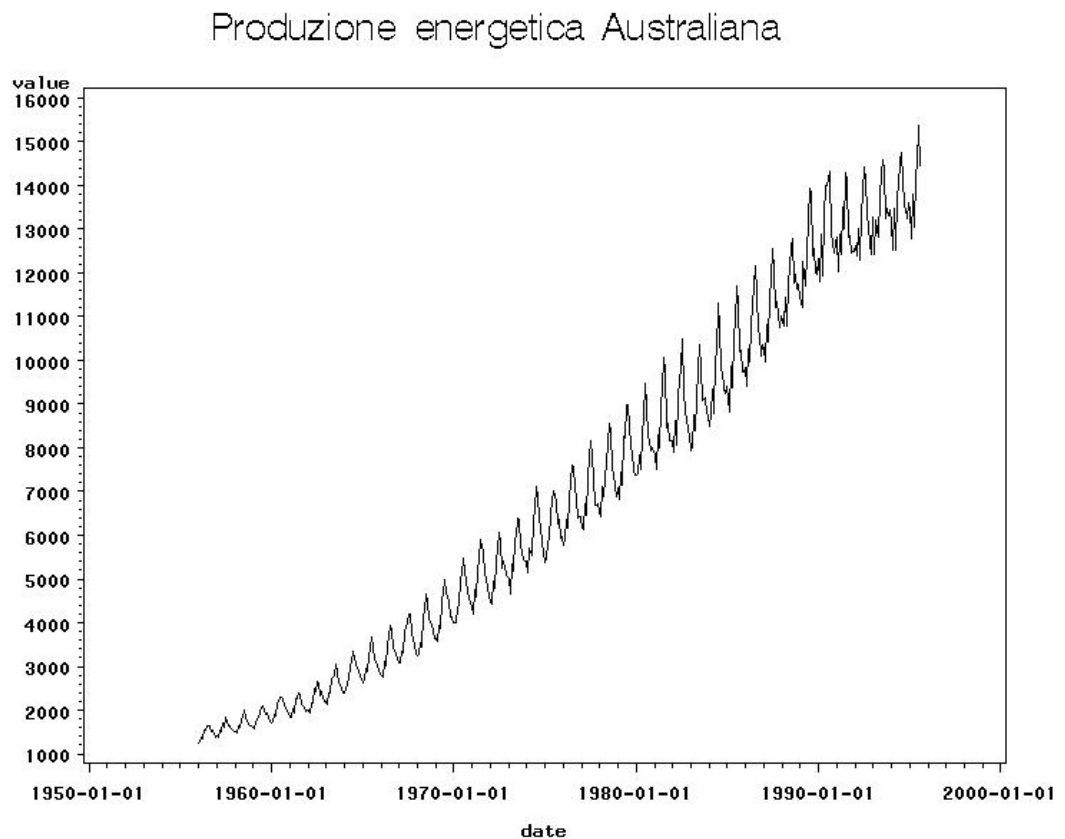
4.5 – Analisi della produzione di elettricità in Australia

Analizziamo ora il secondo dataset, riguardante la produzione mensile di elettricità dell’Australia.

Innanzitutto procediamo ad un'analisi grafica dei dati:

```
proc import out=dati.elect;
    datafile="D:\Home\Tesi\Dati\elect.dat"
    DBMS=TAB replace;
    getnames=YES;
    datarow=2;
run;

proc gplot data=dati.elect;
    plot value*date=1;
    symbol1 line=1 c=black i=join;
    title 'Produzione energetica Australiana';
run;
quit;
```



Al contrario dei dati sul prezzo del petrolio notiamo una netta stagionalità, con i picchi corrispondenti ai periodi invernali (giugno, luglio e agosto), e un andamento in netta crescita, dovuto al progresso della nazione, con conseguente aumento della richiesta di energia.

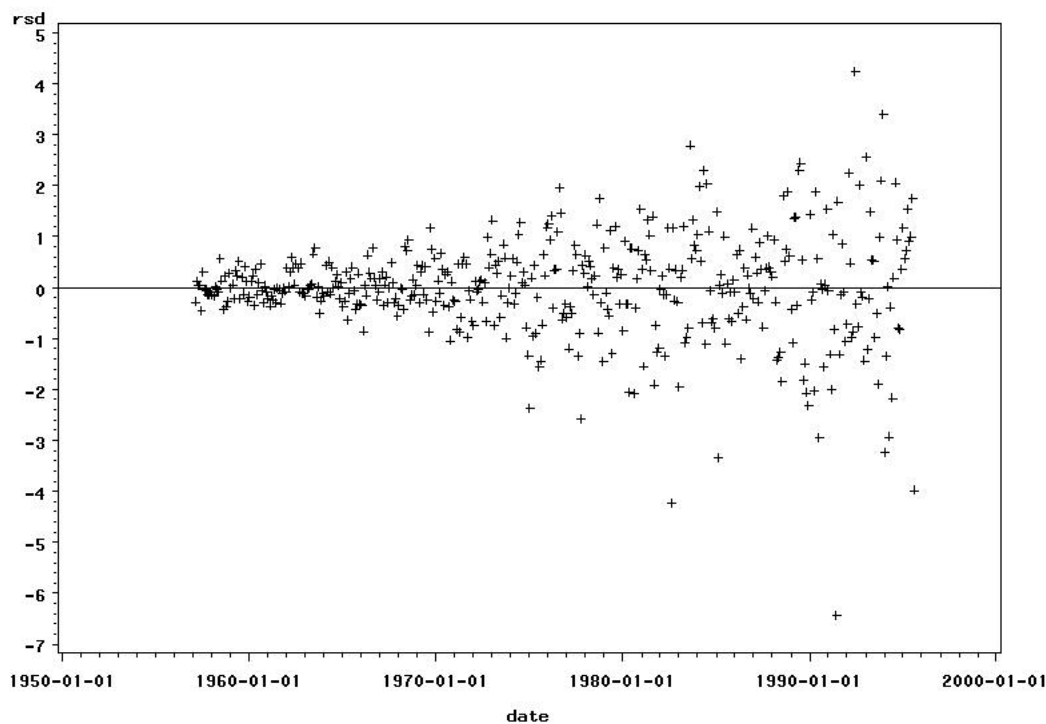
In questa serie notiamo quindi sia un trend marcato, che comporta non stazionarietà in media, sia un aumento della varianza parallelamente al tempo, che porta appunto problemi relativi alla non stazionarietà in varianza.

Il primo importante passo è, quindi, applicare la funzione logaritmo ai nostri dati, per soddisfare il requisito della stazionarietà in varianza.

Questa trasformazione è essenziale per trovare un buon modello. Se infatti saltassimo questo passo, ci ritroveremmo con un modello molto buono per i dati iniziali (con meno varianza) e non buono per i dati successivi (che hanno una maggiore varianza).

Una conseguenza piuttosto evidente sarebbe l'eteroschedasticità nei residui:

Residui della previsione sulla produzione energetica Australiana



Passiamo ora alla ricerca dei parametri del nostro modello. Innanzi tutto l'esperienza pratica ci consiglia, per serie storiche stagionali con trend, come quella che vogliamo analizzare, un utilizzo dell'operatore $\nabla\nabla^{12}$ per eliminare la non stazionarietà in media.

Impostato l'operatore, SAS procede ad una ricerca iterativa dei parametri p e q del modello ARIMA.

```
proc arima data=dati.logelect;
    identify var=value(1,12) esacf p=(0:10) q=(0:10) nlag=48;
run;
quit;
```

Particolarmente interessante risulta, nell'output SAS, la funzione di autocorrelazione inversa:

Inverse Autocorrelations

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
1	0.60192											.	*****										
2	0.41778											.	*****										
3	0.28517											.	*****										
4	0.23268											.	*****										
5	0.25418											.	*****										
6	0.25809											.	*****										
7	0.24682											.	*****										
8	0.18931											.	****										
9	0.18089											.	****										
10	0.24214											.	*****										
11	0.32349											.	*****										
12	0.59361											.	*****										
13	0.36217											.	*****										
14	0.26296											.	*****										
15	0.20394											.	****										
16	0.17736											.	****										
17	0.18543											.	****										
18	0.19730											.	****										
19	0.21703											.	****										
20	0.16578											.	***										
21	0.14293											.	***										
22	0.17548											.	****										
23	0.17330											.	***										
24	0.34125											.	*****										
25	0.17045											.	***										
26	0.12963											.	***										
27	0.12460											.	**										
28	0.13045											.	***										
29	0.13823											.	***										
30	0.14273											.	***										
31	0.14304											.	***										
32	0.09937											.	**										
33	0.07249											.	*										
34	0.06579											.	*										
35	0.03708											.	*										
36	0.12628											.	***										
37	0.04052											.	*										
38	0.03189											.	*										
39	0.04445											.	*										
40	0.04785											.	*										
41	0.06094											.	*										
42	0.06986											.	*										
43	0.07219											.	*										
44	0.05334											.	*										
45	0.03476											.	*										
46	0.02536											.	*										
47	-0.00214											.	.										
48	0.04250											.	*										

Notiamo che i risultati sopra riportati tendono con l'avanzare del tempo ad annullarsi, presentando picchi di evidente autocorrelazione per i multipli di dodici.

Da questa osservazione e dai consueti confronti fra modelli differenti, scegliamo come miglior modello, il SARIMA, ovvero l'ARIMA stagionale. Dopo la centratura, la sua forma analitica è:

$$\nabla\nabla^{12}Log(X_t) = (1 + \theta_1 B)(1 + \theta_{12} B^{12})Z_t,$$

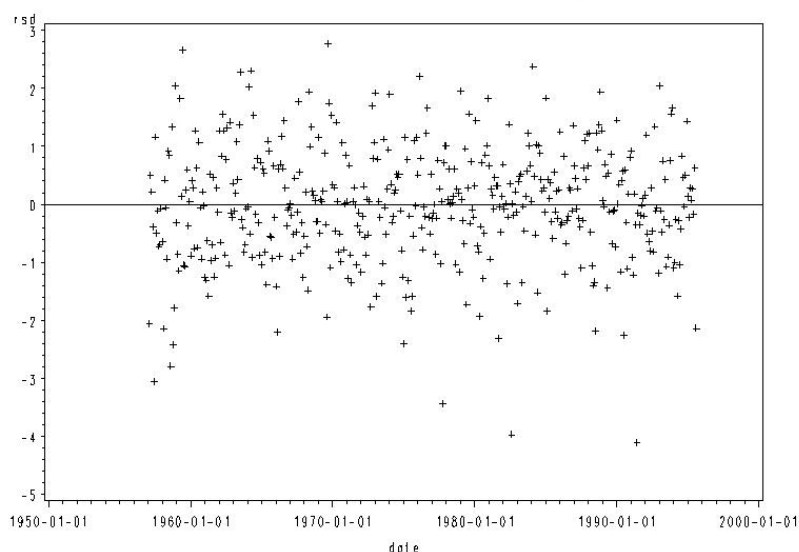
dove $Z_t \sim WN(0, \sigma^2)$, $\hat{\theta}_1 = -0,6726$ e $\hat{\theta}_{12} = -0,6701$.

Verifichiamo dunque con SAS il nostro modello:

```
proc arima data=dati.logelect;
    identify var=value(1,12) esacf p=(0:10) q=(0:10) nlag=48 noprint;
    estimate p=0 q=(1)(12);
    forecast alpha=0.05 lead=48 out=dati.outlogelect
        id=date interval=month;
run;
quit;
```

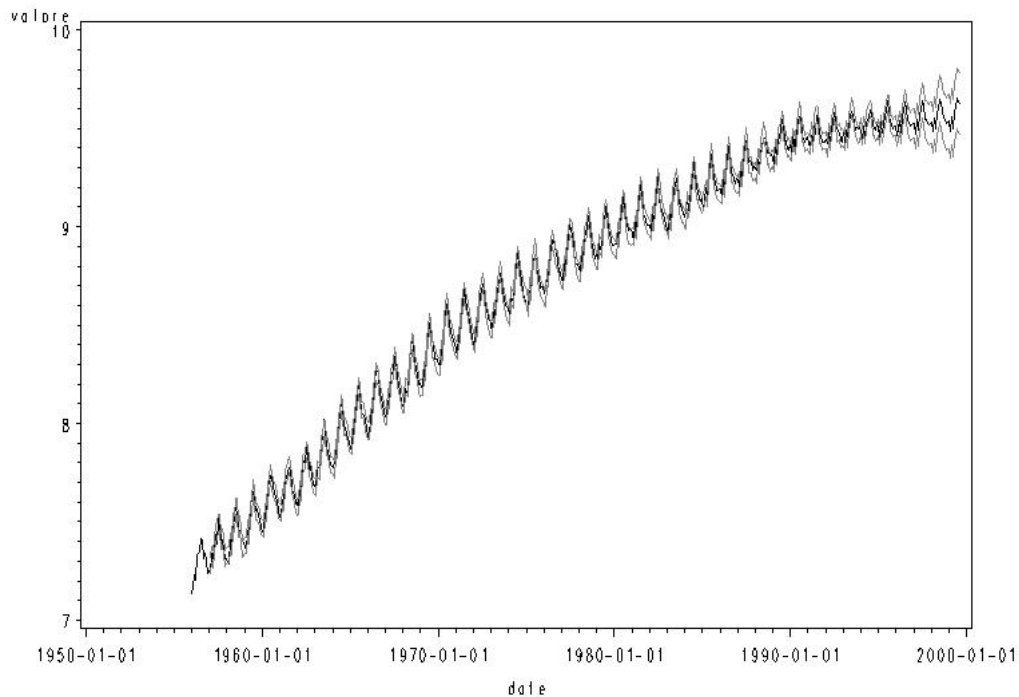
Conditional Least Squares Estimation					
Parameter	Stima	Standard Error	Valore t	Approx Pr > t	Lag
MU	-0.0001874	0.0001118	-1.68	0.0944	0
MA1,1	0.67259	0.03476	19.35	<.0001	1
MA2,1	0.67012	0.03557	18.84	<.0001	12

Residui della previsione sulla produzione energetica Australiana



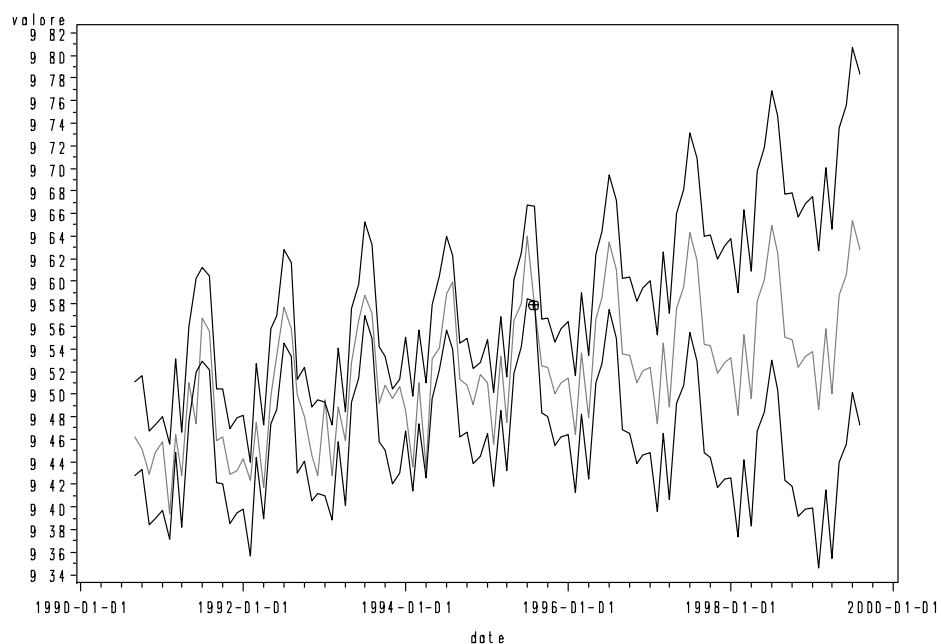
Sia l'analisi dell'output SAS, sia l'analisi grafica dei residui indicano che il modello è soddisfacente. Consideriamo infine come il nostro modello segue graficamente i dati reali e come prevede il futuro:

Intervalli di confidenza sulla produzione energetica Australiana



Il modello trovato sembra che si adatti bene ai dati reali. Per quel che riguarda la previsione (per un periodo di 48 mesi), al contrario del precedente esempio, il modello ci dà un'indicazione molto più precisa e coerente con l'andamento della curva. Nella figura seguente è mostrato nel dettaglio l'andamento degli ultimi anni.

Intervalli di confidenza sulla produzione energetica Australiana



Inoltre l'intervallo di confidenza al 95% si amplia ovviamente all'aumentare dell'orizzonte di previsione, ma in maniera piuttosto contenuta.

Bibliografia

- [1] Domenico Piccolo, Cosimo Vitale (1981), *Metodi statistici per l'analisi economica*, Il Mulino
- [2] Silvia Roncetti (2000), *Modelli autoregressivi per serie storiche*, Tesi di Laurea in Matematica, Università di Genova
- [3] AA.VV. (2003), *SAS[®] Help On Line v.8.02*, SAS Institute
- [4] <http://www.hussmanfunds.com/html/datapage.htm>, *Spot Oil Price: West Texas Intermediate (Jan 1946 – October 2004)*
- [5] <http://www-personal.buseco.monash.edu.au/~hyndman/TSDL>, *Australia monthly production of electricity (Jan 1956 - August 1995)*

Ringraziamenti

È finalmente giunto il momento dei ringraziamenti.

Per primi ringrazio i miei genitori e mio fratello, sempre presenti quando ho bisogno.

Ringrazio tutti i miei amici e parenti, quelli vicini e quelli lontani.

Ringrazio i compagni di corso, che mi hanno regalato due anni intensi e piacevoli, e auguro loro un radioso futuro.

Ringrazio il mio relatore Fabio Rapallo, che mi ha aiutato non poco nella stesura dello scritto.

Ringrazio inoltre tutti i professori dello S.M.I.D. che mi hanno dato l'opportunità e la voglia di continuare a oziare ancora per un paio d'anni.

Se mi sono dimenticato qualcosa o qualcuno mi scuso, ma spero che questo sia solo un primo passo a cui presto seguirà un'altra tesi da scrivere e altri ringraziamenti da fare.