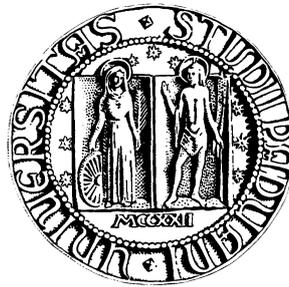


UNIVERSITÀ DEGLI STUDI DI PADOVA

Facoltà di Scienze Statistiche

*Corso di Laurea Specialistica in
Statistica e informatica*



Tesi di Laurea

**Modelli Autoregressivi e carte di controllo
EWMA: effetto della correzione della stima
dei parametri**

Relatore: Prof. Guido Masarotto

Laureando: Paolo Carnà

Anno Accademico 2008/2009

INDICE

INTRODUZIONE.....	3
CAPITOLO 1: MODELLI PER L'ANALISI DELLE SERIE TEMPORALI	5
1.1 <i>Identificazione di un processo stocastico.....</i>	<i>5</i>
1.2 <i>Genesi dei processi ARMA.....</i>	<i>12</i>
1.3 <i>Modelli a Media Mobile (MA).....</i>	<i>14</i>
1.4 <i>Modelli Autoregressivi (AR).....</i>	<i>20</i>
1.5 <i>Modelli a Media Mobile autoregressivi (ARMA)</i>	<i>24</i>
1.6 <i>Stima e previsioni dei modelli ARMA</i>	<i>27</i>
CAPITOLO 2 : METODI STATISTICI PER IL CONTROLLO DI QUALITA': CONCETTI INTRODUTTIVI	33
2.1 <i>Controllo Statistico del Processo (SPC).....</i>	<i>33</i>
2.2 <i>Le carte di controllo</i>	<i>35</i>
2.3 <i>La carta di controllo EWMA: definizione e progettazione</i>	<i>40</i>
2.4 <i>Approccio tramite i residui in presenza di dati autocorrelati</i>	<i>46</i>
CAPITOLO 3 : TECNICHE PER CORREGGERE GLI STIMATORI AUTOREGRESSIVI	48
3.1 <i>Proprietà degli stimatori di un processo ARMA.....</i>	<i>48</i>
3.2 <i>Tecniche per la correzione</i>	<i>51</i>

CAPITOLO 4 : SIMULAZIONE E RISULTATI.....	57
4.1 <i>Scelta dei parametri.....</i>	<i>57</i>
4.2 <i>Interpretazione e discussione dei risultati</i>	<i>60</i>
CONCLUSIONE.....	73
APPENDICE.....	74
BIBLIOGRAFIA.....	83

Introduzione

In un contesto dove si ricerca un miglioramento continuo della qualità nei processi industriali, troverà una sua giustificazione l'applicazione di metodologie dirette al Controllo Statistico del Processo, in quanto rappresentano un primo valido strumento a sostegno dell'attività decisionali ad ogni livello organizzativo per il raggiungimento della qualità. Alla base delle tecniche tradizionali l'ipotesi fondamentale è l'indipendenza delle osservazioni, o incorrelazione, ma in molte situazioni reali, questa condizione non è soddisfatta. I metodi classici per il monitoraggio della qualità di un processo, come le carte di controllo in presenza di dati correlati, forniscono risultati non propriamente accettabili. Di conseguenza negli ultimi anni, sono stati condotti diversi studi al fine di elaborare procedure in grado di adattarsi a processi autocorrelati. Tra questi approcci, uno che si è rivelato essere utile, è stato descrivere direttamente la struttura di correlazione con adeguati modelli di serie temporali, con i quali rimuovere la presenza di autocorrelazione e successivamente applicare la carta di controllo ai residui. L'assunto su cui si basa questa tecnica, è che le stime dei parametri, che descrivono i dati, siano il più vicine possibili al vero valore del parametro. Ovviamente questa situazione porterà ad avere degli errori di stima, e di conseguenza questi ricadranno nelle applicazioni di qualità, come le carte di controllo. Quindi l'obiettivo di questo lavoro è quello di proporre stimatori alternativi che risultino meno distorti degli usuali e di esaminare l'effetto della

correzione per la costruzione di carte di controllo. L'idea di fondo quella di ricorrere alle carte di controllo EWMA (*Exponentially Weighted Moving Average*) basate sui residui per dati autocorrelati, in presenza di incertezza sul modello. In altri ambiti applicativi e' stato dimostrato che lavorare con stimatori meno distorti degli usuali, impiegati per la stima dei parametri di modelli autoregressivi (AR) e di modelli autoregressivi a media mobile (ARMA), porta a risultati superiori. Pertanto lo scopo di questo lavoro consiste nell'analizzare se un eventuale correzione degli stimatori, possa portare a conclusioni migliori anche nel caso delle applicazioni al controllo di qualità. Nel Capitolo 1 saranno trattati modelli per l'analisi delle serie temporali, con i relativi metodi di stima e previsione. Nel Capitolo 2 saranno introdotti concetti fondamentali sulla teoria del Controllo Statistico di Qualità (SPC) e tra questi, in maniera approfondita, le carte di controllo EWMA e le loro proprietà. In seguito nel Capitolo 3 sarà presente un'argomentazione sulle proprietà degli stimatori autoregressivi usuali e saranno proposti stimatori alternativi per la correzione della distorsione. Infine nel Capitolo 4 verranno analizzati e discussi i risultati delle simulazioni svolte, nelle quali gli stimatori tradizionali dei parametri autoregressivi sono stati messi a confronto con quelli proposti al capitolo precedente.

Capitolo 1

Modelli per l'analisi delle serie temporali

1.1 Identificazione di un processo stocastico

I dati a cui vengono applicate le tecniche inferenziali che compongono il bagaglio di uno statistico possono essere di due tipi: *cross-section*, nel caso in cui le osservazioni di cui disponiamo siano relative ad individui diversi, oppure serie storiche, quando ciò che abbiamo sono osservazioni, su una o più grandezze, protratte nel tempo¹. Nel primo caso, per esempio, pensare ad un insieme di N dati osservati come una delle possibili realizzazioni di N variabili casuali indipendenti ed identiche non è un'ipotesi troppo insostenibile. Se rilevo peso e statura di N individui, non c'è ragione di pensare che:

1. le caratteristiche fisiche dell'i-esimo individuo siano in qualche modo connesse a quelle degli altri individui (indipendenza);
2. la relazione fra peso e altezza che vale per l'i-esimo individuo sia diversa da quella che vale per tutti gli altri (identità).

In questi casi, ci serviamo del concetto di realizzazione di una variabile casuale come metafora dell'i-esima osservazione, e l'apparato inferenziale appropriato è uguale a quello standard, in cui l'indipendenza e l'identità ci consentono di dire che

$$f(x_1, \dots, x_N) = \prod_{i=1}^N f(x_i)$$

¹ A dir la verità, un caso intermedio è dato dai cosiddetti dati panel, ma non ce ne occupiamo qui.

cioè che la funzione di densità del nostro campione è semplicemente la produttrice delle funzioni di densità delle singole osservazioni (le quali funzioni sono tutte uguali). Notate che questo tipo di ragionamento è perfettamente appropriato nella maggior parte dei casi in cui i dati da noi osservati provengano da un esperimento controllato, del tipo di quelli che usano i medici o i biologi. Il caso delle serie temporali, tuttavia, presenta una differenza concettuale di base che richiede una estensione dei concetti probabilistici da utilizzare. Questa differenza consiste nel fatto che il tempo ha una direzione, e quindi esiste la storia. In un contesto di serie storiche, infatti, la naturale tendenza di molti fenomeni ad evolversi in modo più o meno regolare porta a pensare che il dato rilevato in un dato istante t sia più simile a quello rilevato all'istante $t-1$ piuttosto che in epoche distanti; si può dire, in un certo senso, che la serie temporale che analizziamo ha "memoria di sé". Questa caratteristica è generalmente indicata col nome di *persistenza*², e differenzia i campioni di serie storiche da quelli cross-section in maniera netta, perché nei primi l'ordine dei dati ha un'importanza fondamentale, mentre nei secondi esso è del tutto irrilevante. Lo strumento che utilizziamo per far fronte all'esigenza di trovare una metafora probabilistica per le serie temporali osservate è il *processo stocastico*. Una definizione di processo stocastico non rigorosa, ma intuitiva, può essere la seguente: un processo stocastico è un vettore aleatorio di dimensione infinita. Un campione di N osservazioni consecutive nel tempo non viene quindi pensato tanto come una realizzazione di N variabili casuali distinte, quanto piuttosto come parte di un'unica realizzazione di un processo stocastico, la cui memoria è data dal grado di connessione fra le variabili casuali che lo compongono. La definizione appena data rende ovvie una serie di proprietà dei processi stocastici piuttosto

² In certi contesti, viene detta anche *isteresi* (o *isteresi*) per indicare circa la stessa cosa.

importanti per il seguito: dato un processo stocastico il cui t-esimo elemento indichiamo con x_t ,

1. è possibile definire concettualmente una funzione di densità per il processo $f(\dots, x_{t-1}, x_t, x_{t+1}, \dots)$;
2. è possibile marginalizzare tale funzione di densità per ogni sottoinsieme delle sue componenti; da questo consegue che sono definite le funzioni di densità marginali per ognuna delle x_t , ma anche per ogni coppia di elementi (x_t, x_{t+1}) e così via; il fatto poi che le osservazioni non siano indipendenti fra loro fa sì che la densità del campione non si può più rappresentare come una semplice produttoria delle marginali;
3. se le funzioni di densità marginali hanno momenti, è possibile dire, ad esempio, che $E(x_t) = \mu_t, Var(x_t) = \sigma_t^2, Cov(x_t, x_{t-1}) = \gamma_{t,t-1}$ e così via;
4. allo stesso modo, è possibile definire funzioni di densità (coi relativi momenti) condizionali.

Le proprietà appena descritte fanno riferimento ai processi stocastici come strutture probabilistiche. Quando però vogliamo utilizzare queste strutture come base per procedure inferenziali, si aprono due problemi:

1. Se quella che osservo (peraltro non nella sua interezza) è una sola realizzazione delle molte possibili, la possibilità logica di fare inferenza sul processo non può essere data per scontata; infatti, non c'è modo di dire quali caratteristiche della serie osservata sono specifiche di quella realizzazione, e quali invece si ripresenterebbero anche osservandone altre.
2. Se anche fosse possibile usare una sola realizzazione per fare inferenza sulle caratteristiche del processo, è necessario che esso sia stabile nel

tempo, cioè che i suoi connotati probabilistici permangano invariati, per lo meno all'interno del mio intervallo di osservazione.

Queste due questioni conducono alla definizione di due proprietà che i processi stocastici possono avere o non avere: *stazionarietà ed ergodicità*. Si parla di processo stocastico stazionario in due sensi: *stazionarietà forte* (anche detta stretta) e *stazionarietà debole*. Per definire la stazionarietà forte, prendiamo in esame un sottoinsieme qualunque delle variabili casuali che compongono il processo; queste non devono necessariamente essere consecutive, ma per aiutare l'intuizione, facciamo finta che lo siano. Consideriamo perciò una "finestra" aperta sul processo di ampiezza k , ossia un sottoinsieme del tipo $W_t^k = (x_t, \dots, x_{t+k-1})$. Questa è naturalmente una variabile casuale a k dimensioni, con una sua funzione di densità che, in generale, può dipendere da t . Se però ciò non accade, allora la distribuzione di W_t^k è uguale a quella di W_{t+1}^k, W_{t+2}^k e così via. Siamo in presenza di stazionarietà forte quando questa invarianza vale per qualsiasi k . In altri termini, quando un processo è stazionario in senso forte le caratteristiche distribuzionali di tutte le marginali rimangono costanti al passare del tempo. La stazionarietà debole, invece, riguarda solo finestre di ampiezza 2: si ha stazionarietà debole se tutte le variabili casuali doppie $W_t^2 = (x_t, x_{t+k})$, hanno momenti primi e secondi costanti nel tempo³; da questo discende che esistono anche tutti i momenti secondi incrociati $E(x_t \cdot x_{t+k})$ con k qualunque, e anch'essi non dipendono da t (anche se possono dipendere da k). A dispetto dei nomi, una definizione non implica l'altra; ad esempio, un processo può essere stazionario in senso forte ma non possedere momenti; viceversa, la costanza nel tempo dei momenti non implica che le varie marginali abbiano la stessa distribuzione. In un caso, tuttavia, le due definizioni coincidono: questo caso, particolarmente importante per le applicazioni pratiche, è quello in cui il

³ È per questo motivo che la stazionarietà debole viene anche definita *stazionarietà in covarianza*.

processo è *gaussiano*, ossia quando la distribuzione congiunta di un qualunque sottoinsieme di elementi del processo è una normale multivariata. Se un processo è gaussiano, stabilire che è stazionario in senso debole equivale a stabilire la stazionarietà in senso forte. Data la pervasività dei processi gaussiani nelle applicazioni ai dati, da un punto di vista operativo si adotta generalmente la definizione di stazionarietà debole, e quando si parla di stazionarietà senza aggettivi, è appunto a questa che ci si riferisce.

Per quanto riguarda l'ergodicità, è una condizione che limita la memoria del processo: un processo non ergodico è un processo che ha caratteristiche di persistenza così accentuate da far sì che un segmento del processo, per quanto lungo, sia insufficiente a dire alcunché sulle sue caratteristiche distributive. In un processo ergodico, al contrario, la memoria del processo è debole su lunghi orizzonti e all'aumentare dell'ampiezza del campione aumenta in modo significativo anche l'informazione in nostro possesso.

Le condizioni sotto le quali un processo stocastico stazionario è ergodico sono troppo complesse per essere descritte brevemente. Sostanzialmente, si può dire che un processo è ergodico se eventi lontani fra loro possono essere considerati indipendenti; osservando il processo per un lasso di tempo abbastanza lungo, è possibile osservare quasi tutte le sottosequenze che il processo è in grado di generare. Il fatto che eventi lontani fra loro nel tempo possano essere considerati indipendenti da un punto di vista pratico è poi spesso sintetizzato nella seguente proprietà dei processi ergodici (che a volte viene usata come definizione di processo ergodico):

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n \text{Cov}(x_t, x_{t-k}) = 0$$

In linea di principio ne consegue che, se un processo è ergodico, è possibile usare le informazioni contenute nel suo svolgimento nel tempo per inferirne le caratteristiche. Esiste un teorema (*teorema ergodico di Slutsky*) che dice che, se

un processo è ergodico, l'osservazione di una sua realizzazione abbastanza lunga è equivalente, ai fini inferenziali, all'osservazione di un gran numero di realizzazioni. Se, ad esempio, un processo ergodico x_t ha valore atteso μ , allora la sua media aritmetica nel tempo è uno stimatore consistente di μ e quindi può essere stimato in modo consistente come se disponessimo di molte realizzazioni del processo anziché di una sola. In linea generale, si può dire che l'inferenza è possibile solo se il processo stocastico che si sta studiando è stazionario ed ergodico. Va detto per altro che, se esistono dei metodi per sottoporre a test l'ipotesi di non stazionarietà, l'ipotesi di ergodicità non è verificabile se si dispone di una sola realizzazione del processo, fosse anche di ampiezza infinita. Nel caso di processi stocastici stazionari, avremo dunque che ogni elemento del processo x_t avrà un valore atteso finito e costante μ e una varianza finita e costante σ^2 . Inoltre, risultano definite tutte le covarianze fra elementi diversi del processo, che saranno pari a

$$\gamma_k = E[(x_t - \mu)(x_{t-k} - \mu)]$$

e che sono note come *autocovarianze*. Si ricordi che la stazionarietà garantisce che queste quantità non sono funzioni di t ; esse sono tuttavia funzioni di k , ed anzi si parla di *funzione di autocovarianza*, intendendo una funzione di k tale per cui $\gamma^{(k)} = \gamma_k$. Va da sé che l'autocovarianza di ordine 0 non è che la varianza e che la definizione è tale per cui $\gamma_k = \gamma_{-k}$, è veritiera. Allo stesso modo è possibile definire le *autocorrelazioni*, che sono date da

$$\rho_0 = \frac{\gamma_k}{\gamma_0} = \frac{\gamma_k}{\sigma^2}$$

Queste quantità, se diverse da 0, costituiscono la memoria del processo, e sono appunto l'elemento che rende i processi stocastici lo strumento teorico adatto a

rappresentare serie temporali caratterizzate da persistenza. Infatti, se $\gamma_1 \neq 0$, allora si ha che

$$f(x_t | x_{t-1}) \neq f(x_t)$$

e di conseguenza

$$E(x_t | x_{t-1}) \neq E(x_t).$$

Potremmo estendere l'insieme di variabili casuali su cui effettuiamo il condizionamento anche a x_{t-2} , x_{t-3} eccetera. Questo insieme di variabili casuali prende a volte il nome di *set informativo* al tempo $t-1$, e viene indicato con I_{t-1} .

1.2 Genesi dei processi ARMA

I processi ARMA costituiscono la famiglia di processi stocastici di gran lunga più utilizzati in statistica e non solo. Questa scelta ha ragioni teoriche e ragioni pratiche, che saranno illustrate nel seguito. Prima di analizzare le caratteristiche principali di tali processi, tuttavia, sono necessarie alcune definizioni di base, che formano l'oggetto delle prossime sezioni. Tanto i processi stocastici che le serie temporali sono, in buona sostanza, sequenze di numeri. Capiterà molto spesso di dover manipolare tali sequenze, e lo si farà per mezzo di appositi operatori. L'*operatore ritardo* (detto anche *operatore Backward*) viene generalmente indicato con la lettera B nella letteratura statistica; è un operatore che si applica a sequenze di numeri, e trasforma una sequenza x_t in un'altra sequenza che ha la caratteristica di avere gli stessi valori di x_t , ma sfalsati di un periodo. Infatti in formule,

$$Bx_t = x_{t-1}.$$

L'applicazione ripetuta n volte di B viene indicata con la scrittura B^n , e quindi si ha $B^n x_t = x_{t-n}$. Per convenzione si pone $B^0 = 1$. L'operatore B è un operatore di tipo lineare nel senso che, se a e b sono costanti, si ha

$$B(ax_t + b) = aBx_t + b = ax_{t-1} + b.$$

La caratteristica più interessante di questo operatore è che le sue proprietà appena enunciate permettono, in molte circostanze, di manipolarlo algebricamente come se fosse un numero. Questo avviene soprattutto quando si considerano polinomi nell'operatore B.

Un'altra definizione da non tralasciare, è quando un processo stocastico è definito *white noise* ("rumore bianco"). Questo tipo di processo è il più semplice che si può immaginare: infatti, è un processo che possiede momenti almeno fino

al secondo ordine; essi sono costanti nel tempo (quindi il processo è stazionario), ma non danno al processo alcuna memoria di sé. La stessa cosa si può dire in modo più formalizzato come segue: un processo white noise, il cui elemento t-esimo indicheremo con ε_t , presenta queste caratteristiche:

$$\begin{aligned}E(\varepsilon_t) &= \mu \\E(\varepsilon_t^2) &= \text{Var}(\varepsilon_t) = \sigma^2 \\ \gamma_k &= 0 \quad \text{per } |k| > 0\end{aligned}$$

Un white noise è quindi, in sostanza, un processo composto di un numero infinito di variabili casuali a media zero e varianza costante; queste variabili casuali, inoltre, sono tutte incorrelate l'una all'altra. A rigore, questo non significa che esse siano indipendenti. Se però si parla di *white noise gaussiano*, ossia di un white noise in cui la distribuzione congiunta di tutte le coppie $(\varepsilon_t, \varepsilon_{t+k})$ sia una normale bivariata, allora sì. Un processo white noise, quindi, è un processo stocastico che non esibisce persistenza. In quanto tale, si potrebbe pensare che sia inadeguato a raggiungere lo scopo di trovare una struttura probabilistica che possa servire da metafora per campioni di serie temporali che, invece, la persistenza ce l'hanno. Il passo in avanti decisivo sta nel considerare cosa succede applicando un polinomio nell'operatore ritardo ad un white noise.

1.3 Modelli a Media Mobile (MA)

Un *processo MA*, o *processo a media mobile* (MA sta appunto per *Moving Average*), è una sequenza di variabili casuali definita tramite la forma

$$y_t = \sum_{i=0}^q \theta_i \cdot \varepsilon_{t-i} = C(B)\varepsilon_t$$

dove $C(B)$ è un polinomio di ordine q nell'operatore ritardo e ε_t è un white noise. Generalmente, e senza perdita di generalità, si pone $C(0)=\theta_0=1$. Se $C(B)$ è un polinomio di grado q , si dice anche che y_t è un processo $MA(q)$. Esaminando i suoi momenti, si vede che quello di ordine primo si scrive:

$$E(y_t) = E \left[\sum_{i=0}^q \theta_i \cdot \varepsilon_{t-i} \right] = \sum_{i=0}^q \theta_i \cdot E[\varepsilon_{t-i}] = 0$$

E quindi un processo MA ha media 0. A prima vista, si potrebbe pensare che questa caratteristica limiti fortemente l'applicabilità di processi MA a situazioni reali, visto che, in genere, non è detto che le serie temporali osservate oscillino intorno al valore 0. Tuttavia, la limitazione è più apparente che reale, visto che per ogni processo x_t per cui $E(x_t)=\mu_t$ si può sempre definire un nuovo processo $y_t=x_t-\mu_t$ a media nulla. Se y_t è stazionario in covarianza, allora basta studiare y_t e poi aggiungere nuovamente la media per avere x_t . Per quanto riguarda la varianza, il fatto che il momento primo sia nullo ci consente di scriverla come il momento secondo, ossia

$$Var(y_t)=E(y_t^2)=E \left[\left(\sum_{i=0}^q \theta_i \cdot \varepsilon_{t-i} \right)^2 \right]$$

Sviluppando il quadrato, possiamo scomporre la somma in due parti:

$$\left(\sum_{i=0}^q \theta_i \cdot \varepsilon_{t-i} \right)^2 = \sum_{i=0}^q \theta_i^2 \varepsilon_{t-i}^2 + \sum_{i=0}^q \sum_{j \neq i} \theta_i \theta_j \varepsilon_{t-i} \varepsilon_{t-j}$$

Dalla proprietà del white noise, che il valore atteso della seconda sommatoria è nullo, cosicché

$$E(y_t^2) = E \left[\sum_{i=0}^q \theta_i^2 \varepsilon_{t-i}^2 \right] = \sum_{i=0}^q \theta_i^2 E(\varepsilon_{t-i}^2) = \sum_{i=0}^q \theta_i^2 \sigma^2 = \sigma^2 \cdot \sum_{i=0}^q \theta_i^2$$

che ha valore finito se $\sum_{i=0}^q \theta_i^2 < +\infty$, cosa sempre vera se q è finito. Infine, con un ragionamento del tutto analogo perveniamo al calcolo delle autocovarianze: l'autocovarianza di ordine k è data da

$$E(y_t y_{t+k}) = E \left[\left(\sum_{i=0}^q \theta_i \varepsilon_{t-i} \right) \left(\sum_{j=0}^q \theta_j \varepsilon_{t-j+k} \right) \right] = \sum_{i=0}^q \theta_i \left(\sum_{j=0}^q \theta_j E(\varepsilon_{t-i} \varepsilon_{t-j+k}) \right)$$

Sfruttando ancora una volta le proprietà del white noise, si ha che $E(\varepsilon_{t-i} \varepsilon_{t-j+k}) = \sigma^2$ se $j = i+k$ e 0 in tutti gli altri casi, cosicché l'espressione precedente si riduce a:

$$V_k = E(y_t y_{t+k}) = \sigma^2 \sum_{i=0}^q \theta_i \theta_{i+k}$$

dove si intende che $\theta_i = 0$ per $i > q$. Si noti che:

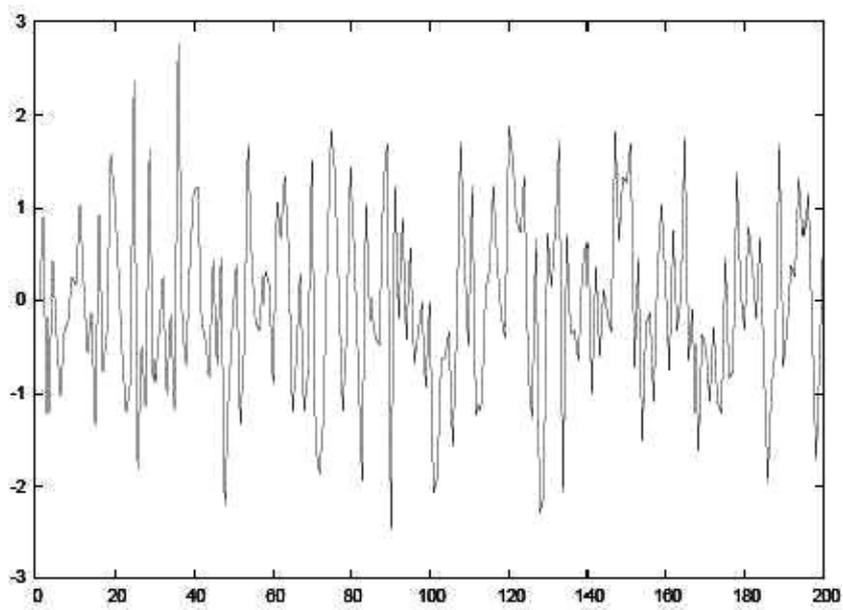
- L'espressione per la varianza è un caso particolare della formula precedente, ponendo $k = 0$;

- per $k > q$, le autocovarianze sono nulle.

Un processo $MA(q)$, quindi, è un processo ottenuto come combinazione di diversi elementi di uno stesso white noise che presenta delle caratteristiche di persistenza tanto più pronunciate quanto più alto è il suo ordine. Quest'ultimo può anche essere infinito; in questo caso, tuttavia, l'esistenza dei momenti secondi (e quindi la stazionarietà) è garantita solo nel caso in cui si verifica $\sum_{i=0}^q \theta_i^2 < +\infty$.

Tanto per avere un'idea più concreta, prendiamo un processo $MA(1)$ di esempio e facciamone un grafico: l'andamento di y_t per diversi valori di q è rappresentato nella figura 1.1. Naturalmente, quando $\theta=0$ il processo è un white noise. Come si vede, al crescere di θ le caratteristiche di persistenza divengono più visibili (la serie temporale si "smussa") e la sua varianza (misurata approssimativamente dall'ordine di grandezza delle ordinate) aumenta.

Figura 1.1: processo $MA(1)$ - $\theta=0$ (white noise)



Se avessimo simulato un processo MA di ordine superiore, la cosa sarebbe stata ancor più evidente. Considerando più a fondo un processo MA(1), si possono fare considerazioni interessanti. Ad esempio, l'autocorrelazione di ordine 1 di un processo MA(1) è data dalla formula

$$\rho_1 = \frac{\theta}{1+\theta^2}.$$

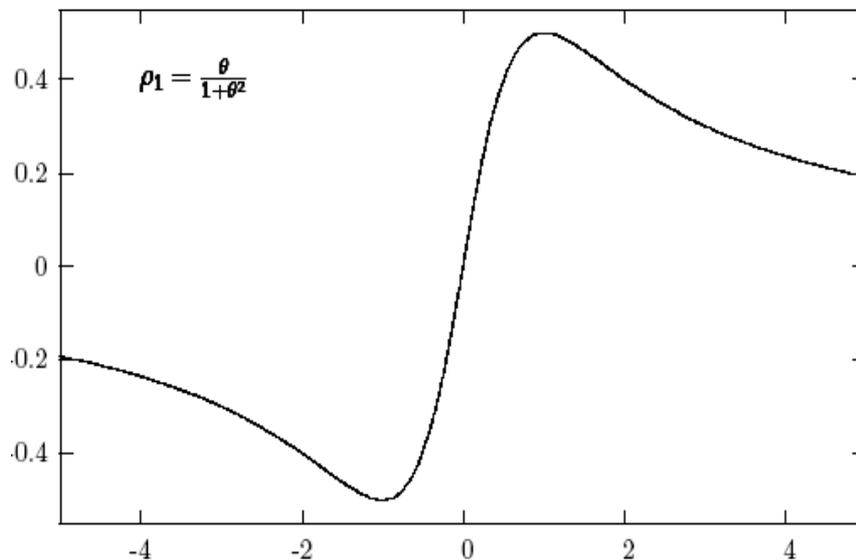
Questa relazione è rappresentata graficamente nella figura 1.2; si può notare che il valore massimo che raggiunge ρ_1 è 0.5, in corrispondenza di $\theta=1$. Un discorso analogo, coi segni cambiati, vale per il punto di minimo. Inoltre, sappiamo dalle considerazioni fatte in precedenza che tutte le autocorrelazioni di ordine maggiore di 1 sono nulle. Poniamoci ora un problema inferenziale: se volessimo rappresentare una certa serie temporale come realizzazione di un processo MA(1), come potremmo utilizzare le statistiche calcolabili sulla serie per ricavare delle stime dei parametri del processo (ovvero, il parametro θ)? Naturalmente, questo procedimento sarebbe sostenibile solo nel caso in cui la nostra serie temporale avesse un correlogramma (il grafico della stima della funzione di autocorrelazione) con valori moderati per l'autocorrelazione di primo ordine e trascurabili per le altre. Se così fosse, si potrebbe anche pensare: se il processo che ha generato i dati è effettivamente un MA(1), allora è stazionario ed ergodico, per cui l'autocorrelazione campionaria converge in probabilità a quella teorica.

In formule:

$$\hat{\rho}_1 \xrightarrow{P} \frac{\theta}{1+\theta^2}$$

poiché questa è una funzione continua di θ , posso invertirla e trovare uno stimatore consistente di θ col metodo dei momenti, ossia trovare quel valore che soddisfa l'equazione

Figura 1.2: MA(1)- Autocorrelazione di primo ordine in funzione di θ



$$\hat{\rho}_1 = \frac{\hat{\theta}}{1 + \hat{\theta}^2}$$

e si vede facilmente che la soluzione è:

$$\hat{\theta} = \frac{1}{2\hat{\rho}_1} \left(1 - \sqrt{1 - 4\hat{\rho}_1^2} \right).$$

In effetti di valori ce ne sarebbero due, ma si è scelto per la soluzione compresa nell'intervallo $[-1,1]$ perché proprio in questo spazio ritroviamo tutti i valori possibili, come ben visibile nel grafico precedente.

Ora, noi sappiamo che le cose non stanno sempre così. È vero che un processo MA di ordine superiore ha autocovarianze più articolate, e quindi si può immaginare che la stessa strategia potrebbe essere percorribile, almeno in teoria, a condizione di specificare un ordine del polinomio $C(B)$ abbastanza alto. Facendo un passo più in là, ci si potrebbe chiedere se la supposizione vale per

qualunque struttura di autocovarianze. La risposta è nel *teorema di Wold*, che essenzialmente dice: qualunque processo stocastico, purché stazionario, possiede una struttura di autocovarianze che è replicabile con un processo a media mobile di ordine infinito. Questo risultato è di importanza enorme: esso ci dice, in sostanza, che qualunque sia la forma “vera” di un processo stocastico stazionario, possiamo sempre rappresentarlo come un processo MA (al limite di ordine infinito). È per questo che, trattando i processi MA, si sta di fatto studiando tutti i processi stazionari possibili, per lo meno per quanto riguarda le loro caratteristiche di media e di covarianza.

1.4 Modelli Autoregressivi (AR)

Un'altra importante classe di processi è data dai processi Autoregressivi. Questi processi forniscono, in un certo senso, una rappresentazione più intuitiva di una serie persistente di quella dei processi MA, poiché l'idea è che il livello della serie temporale al tempo t sia una funzione lineare dei propri valori passati, più un white noise. Il nome deriva appunto dal fatto che un modello AR somiglia molto ad un modello di regressione in cui le variabili esplicative sono i valori passati della variabile dipendente.

$$y_t = \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \varepsilon_t$$

Non è ozioso notare che, in questo contesto, il white noise può essere interpretato in modo analogo al disturbo di un modello di regressione, cioè come la differenza fra y_t e la sua media condizionale; in questo caso, le variabili casuali che costituiscono l'insieme di condizionamento sono semplicemente il passato di y_t . I processi AR sono in un certo senso speculari ai processi MA perché, se un processo MA è un processo definito dall'applicazione di un polinomio nell'operatore B ad un white noise, un processo AR è definito come un processo l'applicazione al quale di un polinomio nell'operatore B produce un white noise. In simboli

$$A(B)y_t = \varepsilon_t$$

dove $A(B)$ è il solito polinomio in b di grado p con $A(0)=1$.

Per familiarizzare con questo tipo di processi, iniziamo col considerare il caso più semplice: quello in cui $p=1$ e il processo può essere scritto

$$y_t = \varphi y_{t-1} + \varepsilon_t \longrightarrow (1 - B\varphi)y_t = \varepsilon_t$$

I momenti di un processo AR(1) possono essere ricavati in diversi modi: uno piuttosto intuitivo è quello di supporre la stazionarietà del processo, e poi derivare le conseguenze da questa ipotesi. Supponiamo quindi che il processo abbia media costante μ . Questa ipotesi implica che:

$$\mu = E(y_t) = \varphi E(y_{t-1}) + E(\varepsilon_t) = \varphi\mu$$

L'espressione precedente può essere vera in due casi: o $\mu = 0$, nel qual caso è vera per qualsiasi valore di φ , oppure nel caso $\varphi = 1$, e allora l'espressione è vera per qualsiasi valore di μ , e la media del processo è indeterminata. In questo secondo caso si dice che il processo presenta una *radice unitaria*, perché il valore di z per cui $A(z)=0$ è appunto 1. Un altro modo di derivare $E(y_t)$ è quello di rappresentare y_t come un processo a media mobile. Per farlo, utilizziamo i risultati riportati precedentemente sulla manipolazione dei polinomi. Se ci limitiamo ai casi in cui $|\varphi| < 1$ (condizione che chiaramente esclude la radice unitaria), si avrà che

$$A(B)^{-1} = (1 - B\varphi)^{-1} = 1 + B\varphi + B^2\varphi^2 + \dots = C(B)$$

e quindi la rappresentazione MA di y_t sarà

$$y_t = (1 + B\varphi + B^2\varphi^2 + \dots) \cdot \varepsilon_t = C(B) \cdot \varepsilon_t$$

cioè un processo MA con $\theta^i = \varphi^i$, che ha media zero. Per quanto riguarda i momenti secondi, procediamo come prima; supponiamo che il white noise abbia varianza pari a σ^2 . Se indichiamo con V la varianza di y_t , e supponiamo che essa esista e sia costante nel tempo, avremo che

$$V = E(y_t^2) = E[(\varphi y_{t-1} + \varepsilon_t)^2] = \varphi^2 V + \sigma^2 + 2\varphi E(y_{t-1}\varepsilon_t)$$

L'ultimo elemento della somma è 0, quindi risulta una combinazione lineare di autocovarianze di un white noise (tutte nulle per definizione).

Se ne deduce che

$$V = \frac{\sigma^2}{1 - \varphi^2}$$

In primo luogo, l'espressione ci dice che solo se $|\varphi| < 1$ ha senso parlare di varianza stabile nel tempo, altrimenti non sarebbe più valida l'ultima uguaglianza. Questa condizione esclude dalla categoria dei processi AR(1) stazionari non solo quelli a radice unitaria, ma anche quelli a radice cosiddetta esplosiva cioè con $|\varphi| > 1$. La seconda considerazione nasce dal confronto di V , che è la varianza non condizionale di y_t , con σ^2 , che è la varianza di $y_t | \mathcal{S}_{t-1}$. V è sempre maggiore di σ^2 , e la differenza è tanto maggiore quanto più φ è vicino a 1: tanto più persistente è il processo, tanto più la sua varianza condizionale al proprio passato sarà minore della sua varianza non condizionale. Vale a dire che la conoscenza del valore di y_{t-1} riduce l'incertezza sul valore di y_t quanto più persistente è la serie. Rimangono da verificare le autocovarianze: l'autocovarianza di ordine 0 è V , che conosciamo già; l'autocovarianza di ordine 1 è data da

$$\gamma_k = E(y_t y_{t-k}) = E[(\varphi y_{t-1} + \varepsilon_t) y_{t-k}] = \varphi \gamma_{k-1}$$

e si deduce che

$$\gamma_k = \varphi^k \frac{\sigma^2}{1 - \varphi^2}$$

Le autocorrelazioni assumono in questo caso una forma molto semplice:

$$\rho_k = \varphi^k$$

Anche in questo caso è possibile dare un'interpretazione intuitiva del risultato: le autocorrelazioni, che sono un indice della memoria del processo, sono tanto più grandi (in valore assoluto), tanto più grande (sempre in valore assoluto) è φ , confermando l'interpretazione di φ come parametro di persistenza. In più, sebbene il limite per k che va all'infinito positivo risulta uguale a zero, ρ_k è sempre diverso da 0. Altri fatti sono che un processo AR(p):

1. ha memoria infinita, ma le autocorrelazioni decrescono al crescere di k in progressione geometrica;
2. nel caso di "intercetta" diversa da 0, ha valore atteso $\mu/A(1)$, dove $A(1)$ è appunto il polinomio $A(z)$ valutato in $z=1$ anziché in $z=B$ come al solito.

L'unico aspetto che vale la pena di sottolineare del caso in cui l'ordine del processo autoregressivo p sia maggiore di 1 è che processi AR(p) possono avere andamenti ciclici. In questo caso, il processo assume un andamento ciclico in cui l'ampiezza delle oscillazioni varia attorno ad un valore medio. Dovrebbe essere evidente che i processi di questo tipo sono i candidati naturali a modellare fenomeni economici caratterizzati da fasi cicliche.

1.5 Modelli Autoregressivi Media Mobile (ARMA)

La classe dei processi ARMA comprende sia i processi AR che i processi MA come caso particolare. Un processo $ARMA(p, q)$ è infatti definito da

$$A(B)y_t = C(B)\varepsilon_t$$

dove p è l'ordine del polinomio $A(B)$ e q è l'ordine del polinomio $C(B)$. Entrambi sono numeri finiti. I processi AR o MA sono quindi casi particolari ($q=0$ e $p=0$ rispettivamente). Se il polinomio $A(B)$ ha tutte le sue radici maggiori di 1 in modulo, allora y_t può anche essere rappresentato in forma MA, ovvero:

$$y_t = A(B)^{-1}C(B)\varepsilon_t = D(B) \cdot \varepsilon_t$$

dove $D(B)$ è un polinomio di ordine infinito se $p > 0$. Tale condizione su $A(B)$ è necessaria e sufficiente affinché il processo sia stazionario. Allo stesso modo, se il polinomio $C(B)$ è invertibile, allora y_t ammette una rappresentazione autoregressiva (di ordine infinito se $q > 0$)

$$A(B)y_t C(B)^{-1} = A(B) \cdot y_t = \varepsilon_t$$

In questo caso, si dice anche che il processo è invertibile. Le caratteristiche dei momenti di un processo $ARMA(p,q)$ possono essere ricavate in modo concettualmente semplice dalla sua rappresentazione in media mobile. L'unica caratteristica che mi pare degna di menzione è che se aggiungiamo un'intercetta, si dimostra che la media del processo è ancora $\mu/A(1)$.

Che senso ha studiare processi ARMA? In linea teorica, nessuna, visto che il teorema di rappresentazione di Wold ci dice che qualunque processo stazionario può essere rappresentato come un processo MA. Da un punto di vista pratico c'è

il problema che la rappresentazione di Wold è, in generale, infinita. Questo non è un problema a livello teorico, ma lo diventa nella pratica: la serie temporale che osserviamo viene infatti pensata come realizzazione di un processo stocastico, i cui parametri sono i coefficienti dei polinomi nell'operatore B che ne determinano le caratteristiche di persistenza (più la varianza del white noise). Se si considera una serie osservata come una realizzazione di un qualche processo stazionario, utilizzare un processo MA per riassumerne le caratteristiche di media e covarianza comporta quindi il problema inferenziale di stimare un numero potenzialmente infinito di parametri. Infatti, se pensiamo che y_t sia rappresentabile in forma MA come

$$y_t = F(B)\varepsilon_t$$

niente ci assicura che il polinomio $F(B)$ non sia di ordine infinito. Si può però pensare di usare un'approssimazione di $F(B)$; in particolare, può darsi che si riescano a trovare due polinomi di ordine finito (e possibilmente basso) $A(B)$ e $C(B)$ tali per cui

$$F(z) \approx \frac{C(z)}{A(z)}$$

Se l'uguaglianza fosse esatta, si potrebbe allora scrivere $A(B)y_t = D(B)\varepsilon_t$ e se quest'ultima vale solo in modo approssimato, allora si avrà:

$$A(B)y_t = D(B)\varepsilon_t^*$$

dove

$$\varepsilon_t^* = \frac{A(B)}{D(B)} \cdot F(B)\varepsilon_t$$

Il processo ε_t^* non è, a rigore, un white noise, ma se le sue autocovarianze non sono troppo grandi, può essere considerato tale a tutti i fini pratici. Si potrebbe dire che considerare ε_t^* un white noise costituisce una metafora dei dati che non è molto più fuorviante di quella basata sulla rappresentazione di Wold e che ha il vantaggio di basarsi su un numero finito di parametri. In pratica, un modello ARMA viene costruito facendo un'ipotesi a priori sui gradi dei due polinomi $A(B)$ e $D(B)$ e poi, una volta stimati i coefficienti dei polinomi, esaminando le autocorrelazioni campionarie della serie. Se queste non sono troppo grandi, non ci sono problemi di sorta a considerare ε_t^* come un white noise. L'esigenza di tener basso il numero dei parametri dei polinomi conduce, in certi casi, a lavorare con dei modelli noti come *ARMA moltiplicativi*, che si usano soprattutto per serie caratterizzate da persistenza stagionale, e che quindi sono anche conosciuti come *ARMA stagionali*, o *SARMA*.

1.6 Stima e previsioni dei modelli ARMA

Fino ad ora abbiamo fatto finta che il processo stocastico che sovrapponiamo ai dati per interpretarli fosse governato da parametri noti. Se questi ultimi noti non sono (e in pratica non lo sono mai), si possono utilizzare delle loro stime. La tecnica di base per la stima dei parametri di un processo ARMA è la *massima verosimiglianza*. Di solito si assume che il processo sia normale, cosicché la forma della funzione di densità delle osservazioni è nota e trattabile. Può essere utile richiamare brevemente cosa si intende per funzione di verosimiglianza. La verosimiglianza è la funzione di densità del campione, calcolata nel punto corrispondente al campione osservato. Essa dipenderà da un vettore y di parametri incogniti, che ne determinano la forma. Massimizzando questa funzione si ottiene la *stima di massima verosimiglianza*. Quando osserviamo una realizzazione di un processo stocastico, la funzione di verosimiglianza non è altro che la funzione di densità congiunta della parte di processo osservata, ossia la funzione di densità marginale del vettore aleatorio (x_1, \dots, x_T) , calcolata nei valori osservati. Se supponiamo che il processo sia gaussiano, la funzione di verosimiglianza non è che la funzione di densità di una normale multivariata:

$$L(\psi) = f(x; \psi) = \left(\frac{1}{2\pi^T} \right) |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (x-k)' \Sigma^{-1} (x-k) \right\}$$

dove x è il vettore (x_1, \dots, x_T) delle T osservazioni; k e Σ sono i suoi momenti primi e secondi, che dipendono da ψ . Ad esempio, l'elemento ij della matrice Σ non è che l'autocovarianza di ordine $|i - j|$ la quale, come sappiamo, è una funzione dei parametri del processo ARMA. È possibile dimostrare che gli stimatori di massima verosimiglianza di processi ARMA gaussiani sono consistenti, asintoticamente normali ed asintoticamente efficienti. Inoltre, sotto condizioni piuttosto blande, le proprietà di consistenza e normalità asintotica vengono conservate anche quando la vera distribuzione del processo non sia

normale (si parla in questo caso di stime di quasi - massima verosimiglianza). Inoltre va detto che, di solito, non si lavora sulla funzione $L(\psi)$, ma sul suo logaritmo ma questo è irrilevante. I problemi principali sono tre:

1. in primo luogo, il sistema di equazioni che risulta uguagliando a 0 il vettore dello score è non lineare, ed in generale non si riesce a risolvere analiticamente, per cui non sono disponibili espressioni che permettano di calcolare gli elementi di $\hat{\psi}$ come semplici funzioni dai dati;
2. non è noto l'ordine dei polinomi $A(B)$ e $C(B)$ adatti a rappresentare in modo adeguato il processo di cui pensiamo x_t sia una realizzazione;
3. per calcolare $L(\psi)$ bisogna conoscere la matrice Σ ; ora, la matrice Σ è una matrice $T \times T$, e quando T è grande (già nell'ordine delle decine, ma i campioni di serie temporali possono assumere dimensioni dell'ordine di decine di migliaia di osservazioni) anche il semplice calcolo dei suoi elementi in funzione di ψ è un problema non da poco, per non parlare del suo determinante o della sua inversa.

Il primo problema meriterebbe una disamina approfondita, ma questa non è la sede. Esistono algoritmi numerici che permettono di trovare il massimo delle funzioni di verosimiglianza per un punto qualunque dello spazio parametrico, senza grandi difficoltà.

Per quanto riguarda il secondo problema, il modo in cui si procede di solito è basato sul fatto che esistono dei criteri, ovvero delle procedure di test, che ci permettono di stabilire se un processo è un white noise o possiede della persistenza; il più diffuso è il cosiddetto *test di Ljung-Box*, che è basato sul fatto che in grandi campioni le autocovarianze campionarie tendono a 0 nel caso di un white noise. Come valori critici vengono considerati quelli di una distribuzione *Chi-quadrato*. Inoltre, come abbiamo già visto, ci sono delle relazioni ben precise fra ordine dei polinomi e autocorrelazioni. Dall'esame delle autocorrelazioni campionarie (che, sotto le ipotesi di ergodicità e stazionarietà sono stimatori consistenti di quelle teoriche) si può fare un'ipotesi di partenza sugli ordini dei

polinomi. Se, ad esempio, si nota che le autocorrelazioni campionarie si interrompono bruscamente al di là di un certo ordine q , si può pensare di usare un modello MA(q), le cui autocorrelazioni teoriche hanno la stessa caratteristica. Se invece le autocorrelazioni digradano dolcemente, forse è meglio un processo AR. Questa fase è nota nella letteratura statistica come *fase di identificazione*. In questa fase, si adoperano a volte anche statistiche note come *autocorrelazioni parziali*. Quest'ultime se considerate di ordine p , si calcolano facendo una regressione di y_t su una costante e $y_{t-1} \dots y_{t-p}$. Il coefficiente associato a y_{t-p} che risulta è l'autocorrelazione parziale di ordine p . Queste grandezze si interrompono bruscamente nel caso di modelli AR puri, e scendono gradualmente nel caso di modelli MA puri. Una volta fatta la stima, si controlla se i residui sono white noise, quasi sempre col test di Ljung-Box o con statistiche equivalenti. Un'altra classe di statistiche che si usano in questo contesto sono i cosiddetti criteri di informazione, come ad esempio quello di *Akaike* (spesso abbreviato in AIC) o quello di *Schwartz* (spesso abbreviato in BIC); l'uso di queste statistiche è motivato con concetti presi dalla teoria dell'informazione, ma non saranno approfonditi in questa sede. Qui basta dire che fra due modelli, quello "migliore" dovrebbe avere un indice AIC o BIC più basso, in quanto tutti questi criteri possono essere scritti nella forma

$$C = -2L(\theta) + c(k, T)$$

dove k è il numero di parametri stimati e T è l'ampiezza campionaria; la funzione $c(k, T)$ è crescente in k , per cui a parità di verosimiglianza viene scelto il modello più parsimonioso. Ad esempio, per il criterio di Schwartz, $c(k, T) = k \log(T)$.

Infine il calcolo della verosimiglianza è il problema più intrigante: bisogna, in sostanza, scrivere la funzione di verosimiglianza con un'espressione alternativa che non richieda il calcolo di matrici di dimensione sproorzionata. Questo

argomento è stato studiato a fondo, ed è bene rinviare ad altre letture per una discussione più esauriente.

Proprio dall'operazione di stima dei parametri di una serie temporale, ci si può allacciare alla teoria delle previsioni di un processo ARMA. La miglior previsione per i valori futuri di y_t si può calcolare sulla base di questo ragionamento: definiamo come *previsore* di y_t una qualche funzione delle variabili contenute nel set informativo \mathfrak{S}_{t-1} . Un previsore, cioè, è una qualche regola che determina la previsione che facciamo su y_t dati i suoi valori precedenti, che supponiamo di conoscere. Se y_t è un processo ARMA (o rappresentabile come tale), un'ipotesi sulla distribuzione di ε_t ci mette in condizione di determinare la distribuzione della variabile casuale $y_t | \mathfrak{S}_{t-1}$. È evidente che questo ci mette in grado anche di determinare la distribuzione condizionale dell'errore di previsione, cioè della variabile

$$e_t = y_t - \hat{y}_t$$

La scelta della distribuzione dell'errore di previsione diventa rilevante se dobbiamo scegliere quale funzione utilizzare come previsore. A rigore, la scelta ottimale dovrebbe essere quella di prendere una funzione che tenga conto dell'errore di previsione e di scegliere un \hat{y}_t tale che la minimizzi. Nella stragrande maggioranza dei casi si prende la media condizionale come previsore. Dato un insieme di osservazioni che vanno da 1 a T, ammettiamo perciò che il miglior previsore di y_{T+1} sia la sua media condizionale al set informativo di cui disponiamo, ossia

$$\hat{y}_{T+1} = E(y_{T+1} | \mathfrak{S}_T)$$

Nel caso di un modello AR puro, la previsione sarà:

$$y_{T-1} = \varphi_1 y_T + \dots + \varphi_p y_{T-p+1}$$

Visto che ancora stiamo sul teorico, qui stiamo assumendo che il set informativo a nostra disposizione si estenda infinitamente all'indietro nel passato, cosa che semplifica molto le cose. Se il nostro set informativo (come accade nella realtà) si interrompe ad una qualche data iniziale, il meccanismo vale ancora per processi stazionari, anche se in modo approssimato. Naturalmente, la valutazione della media condizionale dà un valore puntuale, ma non dice nulla sull'attendibilità della previsione, cioè sulla dispersione dell'errore che ci attendiamo di commettere. In altri termini è necessario valutare anche la varianza dell'errore di previsione. Per fare un esempio, nel caso di un AR(1) avremmo:

$$V(\hat{y}_{T-k}) = \sigma^2 \frac{1 - \varphi^{2k}}{1 - \varphi^2}$$

Può essere interessante notare che la varianza dell'errore di previsione è sempre minore della varianza non condizionale di y_t : questo significa che sfruttare le caratteristiche di persistenza della serie temporale permette di rendere meno incerto il suo comportamento futuro. Peraltro, per k che tende all'infinito, le due varianze tendono a coincidere, e questo avviene perché nei processi AR(1) stazionari la persistenza ha sempre un carattere di breve periodo. La conoscenza dello stato del sistema non è informativa sul futuro remoto del sistema stesso più di quanto non lo sia la sua distribuzione non condizionale: per k abbastanza grande, y_t e y_{t+k} sono virtualmente incorrelate (e quindi, se gaussiane, virtualmente indipendenti).

Nel caso più generale di processi ARMA, le previsioni si possono fare applicando ancora lo stesso concetto. In particolare, si noti che, se il set informativo non ha limite temporale inferiore, allora esso comprende non solo tutti i valori passati di y_t , ma anche quelli di ε_t . Il fatto che poi il set informativo a nostra disposizione

non sia infinito rappresenta solo un problema di minore entità. Se infatti noi abbiamo solo osservazioni nell'arco di tempo $[0, \dots, T]$, una soluzione molto comoda è quella di estendere il nostro set informativo all'indietro usando i valori medi non condizionali di $y-1$, $y-2$, eccetera. Se il processo è stazionario ed ergodico, al crescere del campione non c'è differenza. Il calcolo esatto, volendo, si può fare. Ci sono molti modi, ma il più comune, anche perché facilmente automatizzabile, è quello di usare un filtro che si chiama *filtro di Kalman*, che non sarà trattato in questo lavoro. In conclusione va fatta un'osservazione. Ci si potrebbe legittimamente chiedere quale sia il valore pratico delle previsioni fatte in questo modo; in effetti, si ha sempre a che fare con serie temporali a cui associamo processi ARMA solo come rappresentazione stilizzata e approssimata. In altri termini, le caratteristiche storiche di persistenza della serie vengono sintetizzate giocando a far finta che la serie che osserviamo sia una realizzazione di un processo ARMA che ha proprio quelle caratteristiche. Non c'è nessun motivo logico, però, per cui una approssimazione che andava bene per il passato continui ad andare bene per il futuro. Per considerare attendibile una previsione di una serie storica fatta in questo modo, è necessario assumere, più o meno implicitamente, che l'insieme di circostanze che hanno fino ad oggi congiurato a far sì che quel certo processo fosse una buona approssimazione dell'andamento di quella certa serie storica continuino a valere per l'orizzonte temporale che ci interessa prevedere. Questa condizione è spesso verosimile quando la serie temporale è una descrizione di un fenomeno fisico ragionevolmente stabile (ad esempio, la temperatura rilevata giornalmente nello stesso luogo, alla stessa ora). Tuttavia, nel caso di altri fenomeni questa può essere un'ipotesi piuttosto coraggiosa, in quanto la catena causale di eventi che concorrono a determinare il valore della serie in un dato momento è probabilmente più instabile. Per meglio dire, la previsione di un modello ARMA va presa per buona come previsione condizionale ad uno scenario.

Capitolo 2

Metodi statistici per il controllo della qualità: concetti introduttivi

2.1 Controllo Statistico del Processo (SPC)

In un quadro di riferimento dove è necessario un continuo miglioramento della qualità in ambiti aziendali e non solo, trova una sua giustificazione l'applicazione di metodologie dirette al Controllo Statistico del Processo. Queste tecniche risultano fondamentali in quanto sono esse a rappresentare un primo valido strumento a sostegno dell'attività decisionale ad ogni livello organizzativo per il raggiungimento della qualità. Il Controllo Statistico del Processo (SPC) può essere definito come una metodologia che, in riferimento ad una determinata attività, operazione, fase o processo caratterizzato da ripetitività, fa ricorso a tecniche statistiche al fine di definire, analizzare e verificare le condizioni che determinano la variabilità dell'oggetto di analisi. Gli studi sull'SPC non sono certo temi nuovi; già nel 1924, infatti, Shewart iniziò a sviluppare un approccio statistico al controllo di qualità, rilevando che il concetto di variabilità riferito ai fenomeni naturali era ugualmente adeguato all'analisi e alla descrizione dei processi produttivi. Con il contributo della statistica inferenziale e della statistica descrittiva, arrivò allora alla descrizione sintetica di fenomeni più ampi da impiegare come modelli di supporto. Nacque così la *carta di controllo Shewart*, uno degli strumenti statistici più impiegati nell'analisi dei processi produttivi. Da allora i passi avanti compiuti sul tema sono stati molti. Primo fra tutti il riconoscimento circa la validità di questi strumenti e un loro più vasto impiego.

Tornando all'analisi di un processo produttivo, questo opera nel tempo realizzando una serie di prodotti che possono essere considerati elementi della popolazione di pezzi che il processo può produrre. E' tuttavia presente una variabilità delle prestazioni del processo in quanto nessun pezzo prodotto è uguale ad un altro. Ovvero, misurando una medesima caratteristica X (che rappresenta l'elemento di qualità che interessa) su ogni prodotto, si osserverà una certa variabilità della stessa. La presenza di variabilità giustifica pienamente l'approccio statistico. E infatti la modalità o valore della caratteristica X , rilevata sul singolo prodotto, viene vista come la determinazione di una variabile casuale con una data distribuzione di probabilità. In particolare, si deve tenere conto che:

1. in ogni punto nel tempo (ricordiamo che una caratteristica del processo di produzione è la dimensione temporale), la grandezza X può essere descritta da un particolare modello distributivo. Qui faremo riferimento ad un *modello parametrico*: conoscendo i valori dei parametri siamo in grado di identificare perfettamente la distribuzione;
2. lo specifico valore osservato su un prodotto può essere considerato come un valore generato dal quel particolare modello e cioè può essere visto come un campione casuale semplice di 1 unità, estratto dalla popolazione caratterizzata da quel modello distributivo;
3. la distribuzione di X può cambiare nel tempo ovvero possono cambiare nel tempo i valori dei parametri distributivi. Quando intervengono tali modifiche significa che ci sono state variazioni sistematiche ovvero la popolazione è cambiata.

Dopo queste doverose premesse, una descrizione delle principali metodologie statistiche utilizzate per il controllo e il miglioramento della qualità di un processo, sarebbe opportuna ma questa non è la sede. L'attenzione si focalizzerà sulla descrizione e sul funzionamento delle carte di controllo, in modo

particolare della carta di controllo EWMA, impiegata nell'ambito del lavoro di tesi.

2.2 Le carte di controllo

Al fine di ottenere livelli di qualità accettabili può essere determinante intraprendere un'azione di monitoraggio della variabilità del processo produttivo. Una eccessiva variabilità comporterebbe una non rispondenza del prodotto alle sue caratteristiche funzionali. In ogni processo produttivo è presente una variabilità intrinseca che non dipende da cause esterne, ovvero una variabilità naturale originata da una serie di fluttuazioni interne al processo, risultato di numerose piccole cause che operano casualmente (dette *cause comuni o casuali*). Tali cause risiedono nel sistema di produzione e non possono essere attribuite ad esempio a macchine, dipendenti o fornitori particolari; in questo caso la causa ultima risiede e va ricercata nel sistema di produzione, che deve essere modificato, e non in un aspetto specifico del processo. Un processo produttivo la cui fonte di variabilità è imputabile esclusivamente a questo tipo di cause è un processo prevedibile, che può essere descritto mediante leggi statistiche. Si parla in questo caso di processo "sotto controllo statistico". Sulla variabilità del processo possono però intervenire fattori esterni che ne alterano la variabilità naturale e generano una variabilità non prevedibile che disturba il funzionamento del processo. Tali fattori, denominati *cause speciali*, determinano la quota eccezionale di variabilità e rappresentano grosse fluttuazioni nei dati, che non sono imputabili al processo oggetto di analisi. Queste fluttuazioni sono il risultato di cambiamenti all'interno della serie, che possono indicare il verificarsi di problemi oppure, al contrario, l'insorgere di novità interessanti da esplorare.

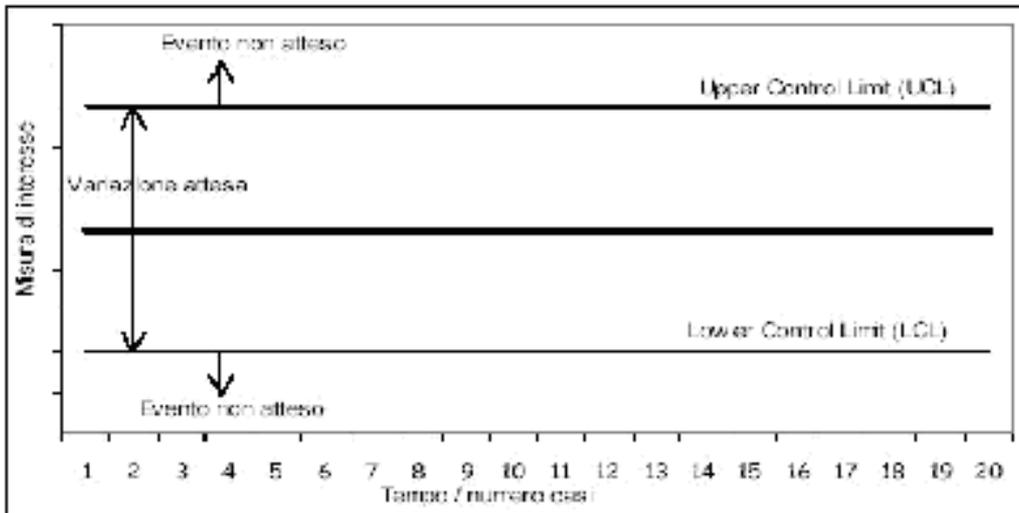
Un processo la cui variabilità risente oltre che di cause comuni anche di cause speciali di variazione ha un andamento imprevedibile, per cui si parlerà di processo *fuori controllo statistico*. Saper distinguere fra le due cause di variabilità è essenziale, in quanto mentre le cause speciali di variazione possono essere corrette (se necessario) senza modificare il processo, le cause ordinarie di variazione possono essere ridotte solo cambiando il processo medesimo.

La carta di controllo, supportando l'analista nel riconoscimento della causa di variazione, consente di individuare un processo fuori controllo consentendo di evitare due errori tipici. Il primo consiste nell'interpretare una causa comune di variazione come una causa straordinaria; in questo caso si potrebbe esercitare un'azione correttiva eccessiva che può a sua volta aumentare la variabilità del processo. Il secondo è l'errore speculare, si commette quando una variazione straordinaria viene trattata come una variazione comune. In questo caso, si rischia di non intervenire prontamente e adeguatamente per "correggere" il sistema.

Le carte di controllo rappresentano uno dei metodi più utilizzati per il controllo statistico di produzione. Queste sono sostanzialmente rappresentazioni grafiche di un processo nel tempo che, basandosi su teorie statistiche, rimangono di comprensibile interpretazione e utilizzo. In letteratura esistono diversi tipi di carte di controllo, la cui forma generale è riportata in figura 2.1.

Le tre linee orizzontali continue chiamate *linea centrale (CL)*, *limite superiore di controllo (UCL)* e *limite inferiore di controllo (LCL)* definiscono la tendenza centrale e un intervallo di variazione naturale per i valori riportati sul grafico. I limiti inferiori e superiori sono calcolati in base a una distribuzione di frequenza teorica che cambia in funzione del tipo di dati che vengono analizzati. Tipicamente una carta di controllo stabilisce dei limiti che si collocano a ± 3 scarti quadratici medi dalla misura statistica di interesse (*metodo dei 3-sigma*).

Figura 2.1 – Carta di controllo generale



Quindi in generale per costruire una carta possiamo seguire la seguente semplice regola (indicando con W una statistica campionaria generica) :

$$CL = E[W]$$

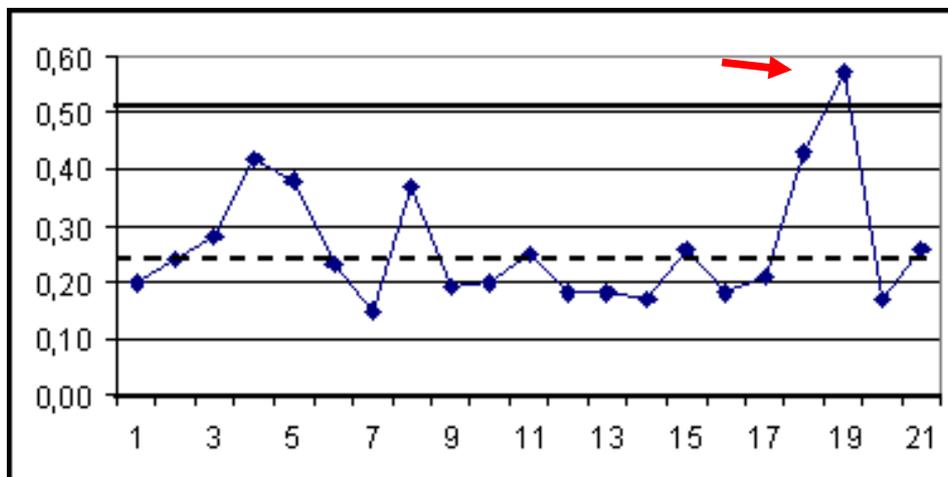
$$UCL = E[W] + 3 (Var[W])^{1/2}$$

$$LCL = E[W] - 3 (Var[W])^{1/2}$$

Data una distribuzione di frequenza teorica di riferimento, l'interpretazione dei valori esterni alle linee di controllo inferiore e superiore è simile a quella di un generico test statistico di ipotesi, indicando come statisticamente significativi i valori che sono *fuori controllo*, testimonianza questa di un processo non omogeneo o comunque di un processo che produce un output sensibilmente diverso da quello di riferimento. Dunque, indipendentemente dal tipo di carta di controllo utilizzata, la lettura può considerarsi sempre la stessa. Una volta definiti i limiti di controllo, riportando i dati all'interno del grafico, la carta ci consente di individuare eventuali andamenti sistematici dei valori che rappresentano il processo nel tempo e di stabilire se ciascun punto cade

all'interno o all'esterno dei limiti imposti. In questo modo si individua immediatamente un processo fuori controllo. Nella pratica raramente vengono considerate singole osservazioni, il più delle volte i valori riportati nelle carte di controllo rappresentano a loro volta il risultato di una stima campionaria (pertanto il singolo valore riportato nella carta di controllo non rappresenta una singola misura, ma, ad esempio, la media di una serie di misure eseguite campionando vari elementi dallo stesso lotto). I vari campioni utilizzati, vengono chiamati *sottogruppi* e si presume che al loro interno presentino la sola variabilità casuale, mentre le cause speciali di variazione (se presenti) possono determinare solamente la variabilità fra diversi sottogruppi. Il grafico 2.2 mostra un tipico esempio di carta di controllo in cui al diciannovesimo periodo si ha un fuori controllo; questo è sintomatico di un processo fuori controllo statistico sulla cui variabilità intervengono cause speciali di variazione.

Figura 2.2 – Identificazione di un fuori controllo



La presenza di tutti i punti della carta all'interno dei limiti di controllo è solo una condizione necessaria ma non sufficiente per poter dire che un processo è *in*

controllo. Si possono individuare sequenze temporali particolari che evidenziano che qualcosa di anomalo sta intervenendo nel processo.

Ad esempio, si può immaginare che un processo abbia la maggioranza dei punti collocata al di sotto del limite centrale, questo induce a pensare ad una anomalia nel processo. Situazioni analoghe possono presentarsi anche nel caso in cui si evidenzia un andamento ciclico dei dati dovuto sicuramente a delle cause ricorrenti o nel caso in cui i punti del tracciato tendano a cadere vicino ai due limiti di controllo con assenza di punti nel centro. Un caso ancora più evidente si verifica quando si individua un trend di crescita, o di decrescita, indicando una anomalia di processo (si parla di *tendenza*). L'individuazione di eventuali anomalie nel tracciato di dati di una carta di controllo non è comunque compito semplice e necessita di una buona conoscenza del processo in esame.

Comunque l'efficienza delle carte di controllo viene valutata tramite la rapidità con cui vengono segnalati eventuali fuori controllo e dal numero di segnalazioni di falsi allarmi. La valutazione dell'efficienza di una carta avviene in termini *Run Length* (RL) e del suo valore atteso, detto *Average Run Length* (ARL). La Run Length è una variabile casuale che rappresenta il primo istante temporale in corrispondenza del quale la statistica di controllo fuoriesce dai limiti di controllo stabiliti. In una situazione ideale fintanto che il processo rimane in controllo, l'ARL della carta dovrebbe essere estremamente grande, mentre al minimo cambiamento la segnalazione dovrebbe avvenire all'istante. In pratica, certe anomalie di modeste entità, vengono indicate con un certo ritardo o non vengono affatto segnalate (*fuori controllo non segnalati*), mentre valori assolutamente accettabili causano a volte *falsi allarmi*.

2.3 La carta di controllo EWMA: definizione, proprietà e progettazione

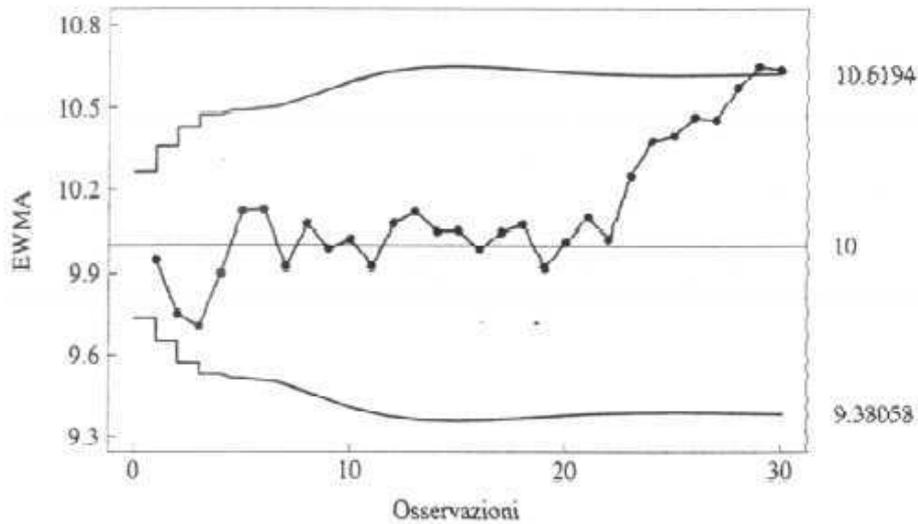
Le prime carte di controllo comunemente utilizzate presero il nome da Shewhart, il quale per primo utilizzò i dati a sua disposizione formulando diversi modelli grafici che si differenziano in base alle caratteristiche stesse dei dati e che sostanzialmente si dividono in due gruppi: *per variabili*, che si basano su misure quantitative e *per attributi* che invece, utilizzano delle misure qualitative. Si ricorda, per esempio, che le carte Shewhart e le carte di controllo per variabili assumono come ipotesi iniziale per la loro applicazione che i dati provengano da una distribuzione Normale e considerano esclusivamente informazioni rilevate all'ultimo istante d'osservazione, risultando così insensibili a piccole modifiche del livello del processo. In questi casi in letteratura viene suggerito l'utilizzo di *carte CUSUM* (Cumulative Sum) ma soprattutto di *carte EWMA* (Exponentially Weighted Moving Average) direttamente sui dati autocorrelati. Questa tipologia di carta di controllo *a media mobile pesata esponenzialmente*, risulta essere una valida alternativa alle carte citate in precedenza poiché tiene conto di salti di livello di entità limitata, mantenendo prestazioni equivalenti.

In sintesi, le carte CUSUM sono più sensibili rispetto ai piccoli slittamenti della media, ma ovviamente perdono molta sensibilità rispetto alle sue brusche variazioni. Differenza fondamentale tra questa carta e quella Shewhart, è che la posizione del punto della carta Shewhart dipende solo dall'ultima misura, mentre nella CUSUM dipende dai valori di tutte le misurazioni fatte in precedenza. Possiamo dare una diversa interpretazione alle due procedure. In pratica esse equivalgono a realizzare una media pesata dei risultati delle misurazioni fatte; nella CUSUM, a tutte è assegnato lo stesso peso, mentre nella carta di controllo Shewhart è assegnato peso pari a uno all'ultima osservazione e peso pari a zero a tutte le altre. Per ottenere una carta che metta insieme tutti i vantaggi delle due carte di controllo precedenti, si potrebbe usare una media mobile, che tuttavia ha l'inconveniente di produrre risultati che tendono ad oscillare. Migliori risultati invece, si ottengono con la media mobile pesata

esponenzialmente (carta EWMA) che assegna ai dati pesi tanto più piccoli quanto più sono stati rilevati in istanti distanti dall'attuale. Rispetto alle altre carte di controllo descritte, questa ha il vantaggio di consentire un pronto intervento sul sistema sia in presenza di piccoli o di grandi slittamenti della media. Inoltre le carte di controllo EWMA risultano avere proprietà di robustezza alla Normalità. Infatti vengono considerate che le assunzioni di base, per un disciplinato uso delle carte di controllo, stabiliscono che i dati generati da un processo in stato di controllo sono indipendenti e si distribuiscono come una Normale di media μ e deviazione standard σ , considerate costanti e ignote. Una situazione di fuori controllo comporta un cambiamento di uno dei due parametri, oppure di entrambi. Quando queste condizioni sono soddisfatte, si possono applicare le carte di controllo convenzionali, le cui proprietà statistiche costituiscono un agevole guida all'interpretazione dell'andamento del processo in analisi. Ma si è dimostrato che anche in presenza di violazioni dell'assunto di normalità della distribuzione, la carta di controllo EWMA continua a funzionare in modo soddisfacente. Il requisito più importante è quello dell'indipendenza delle osservazioni, in mancanza della quale la carta non funzionerebbe in modo corretto, producendo un'elevata quota di falsi allarmi. Purtroppo le assunzioni di indipendenza e di incorrelazione non sono soddisfatte in molte situazioni reali. In particolare la carta di controllo EWMA (esempio figura 2.3) fu introdotta da Roberts (1959) e può essere adoperata in due diversi modi: come strumento per monitorare il processo produttivo e come previsore per le successive osservazioni del processo stesso.

Si assuma che le osservazioni della caratteristica di qualità di interesse, denotata con X_t , siano indipendenti ed identicamente distribuite secondo una Normale di media μ_t e deviazione standard, σ_x dove il valore μ_t indica il valore della media delle osservazioni, la quale può subire dei cambiamenti nel tempo.

Figura 2.3 – Esempio di carta di controllo EWMA



Con W_t indicheremo il valore della statistica EWMA al tempo t , che si può definire come:

$$W_t = (1 - \lambda)W_{t-1} + \lambda y_t$$

dove λ è la costante di liscio esponenziale appartenente all'intervallo $0 < \lambda < 1$ che determina quanto peso assegnare alle osservazioni passate nel calcolo di W_t . Più il suo valore sarà piccolo, più memoria avrà la statistica. L'equazione precedente può essere scritta come:

$$W_t = \lambda \sum_{i=0}^{t-1} (1 - \lambda)^i y_{t-i} + (1 - \lambda)^t W_0$$

dove W_0 è pari al valore target della statistica. Assumendo che $W_0 = \mu$ è facile dimostrare che dall'equazione antecedente risulta $E[W_t] = \mu$, mentre la varianza è pari a:

$$\sigma_{W_t}^2 = \sigma_x^2 \left(\frac{\lambda}{2-\lambda} \right) [1 - (1-\lambda)^{2t}]$$

Di conseguenza, la linea centrale e i limiti di controllo per la carta EWMA vanno posti ai seguenti valori:

$$UCL = \mu_0 + h\sigma \sqrt{\frac{\lambda}{2-\lambda} [1 - (1-\lambda)^{2t}]}$$

$$CL = \mu_0$$

$$LCL = \mu_0 - h\sigma \sqrt{\frac{\lambda}{2-\lambda} [1 - (1-\lambda)^{2t}]}$$

dove h è una costante che determina l'ampiezza dei limiti di controllo. Dal momento che per t che tende all'infinito, la quantità

$$1 - (1-\lambda)^{2t} \longrightarrow 1$$

si otterranno i risultati asintotici. Nello specifico la varianza asintotica sarà:

$$\sigma_{W_t}^2 = \sigma_x^2 \left(\frac{\lambda}{2-\lambda} \right)$$

e i limiti di controllo asintotici risultano essere:

$$UCL = \mu_0 + h\sigma \sqrt{\frac{\lambda}{2-\lambda}}$$

$$LCL = \mu_0 - h\sigma \sqrt{\frac{\lambda}{2-\lambda}}$$

Tuttavia è raccomandabile utilizzare le formule esatte per valori piccoli di t , al fine di ottenere buone prestazioni da parte della carta di controllo. Come accennato in precedenza, gli andamenti non casuali all'interno del processo vengono segnalati solitamente con la fuoriuscita della statistica W_t dai limiti di controllo. Quindi in presenza di una carta di controllo EWMA, andamenti non casuali possono essere riconosciuti anche studiando la distribuzione della statistica all'interno dei limiti di controllo. Per ottenere una carta di controllo EWMA che sia in grado di raggiungere determinate performance, è necessario individuare valori ottimali per la coppia di parametri, λ e h . Come spiegato nel paragrafo precedente, i criteri per valutare l'efficienza di una carta sono basati maggiormente sulla distribuzione della Run Length (RL) e sul suo valore atteso ARL. L'obiettivo sarà dunque, costruire una carta di controllo che segnali il minor numero di falsi allarmi e che riesca ad identificare il prima possibile un vero cambiamento del processo. Di conseguenza la carta di controllo dovrà essere caratterizzata da limiti di controllo ottimali in termini di ARL, in controllo e fuori controllo. Per un ARL in controllo fissato, i parametri vengono determinati in modo che l'ARL fuori controllo risulti il più piccolo possibile per un cambiamento di grandezza fissata. Pertanto per una carta di controllo EWMA, ciò consiste nel raggiungere valori ottimali per le costanti λ e h . Il numero di combinazioni di λ e h per ottenere una carta EWMA con un determinato ARL in controllo sono infinite, ma è necessario determinare quella combinazione in grado di produrre il minor ARL fuori controllo quando un cambiamento è effettivamente avvenuto. L'approccio generale per il disegno di una carta EWMA è il seguente:

3. stabilire l'ARL in controllo della carta;
4. determinare il cambiamento minimo (detto *shift*) che la carta deve individuare;
5. trovare la costante di lisciamento, λ^* , utilizzando delle tavole appropriate [Crowder (1987b, 1989) o di Lucas e Saccucci (1990)];

6. dato λ^* , trovare il valore di h^* che soddisfa il vincolo sull'ARL in controllo;

Molti studi hanno approfondito le proprietà della carta di controllo EWMA in termini di ARL. Si ricordano i lavori di Crowder (1987-1989) e Lucas e Saccucci (1990); in essi vengono forniti tabelle dei valori di ARL e grafici per diversi valori di λ e h . Ad esempio nella tabella 2.4 sono indicate prestazioni ARL di alcuni schemi tipici di controllo. In generale, si è dimostrato che valori di λ compresi nell'intervallo [0.05 - 0.25] forniscono prestazioni accettabili nel caso si è interessati ad individuare piccoli salti di media. Mentre per $h=3$ (che corrisponde agli usuali limiti di controllo a 3-sigma) è una scelta ragionevole per valori grandi del parametro di lisciamiento.

Tabella 2.4 - Valori di ARL per alcune carte EWMA (adattato da Lucas e Saccucci)

Salto di media	$h = 3.054$	2.998	2.962	2.814	2.615
	$\lambda = 0.4$	0.25	0.20	0.10	0.05
0	500	500	500	500	500
0.25	224	170	150	106	84.1
0.50	71.2	48.2	41.8	31.3	28.8
0.75	28.4	20.1	18.2	15.9	16.4
1.00	14.3	11.1	10.5	10.3	11.4
1.50	5.9	5.5	5.5	6.1	7.1
2.00	3.5	3.6	3.7	4.4	5.2
2.50	2.5	2.7	2.9	3.4	4.2
3.00	2.0	2.3	2.4	2.9	3.5
4.00	1.4	1.7	1.9	2.2	2.7

2.4 Approccio tramite i residui in presenza di dati autocorrelati

Come accennato al paragrafo precedente, in molte circostanze reali i dati che noi trattiamo, risultano essere dipendenti correlati fra loro. Si può pensare ad un approccio per serie temporali ma a volte può risultare difficoltoso poiché le carte di controllo vengono applicate a molte variabili del processo, cosicché lo sviluppo di modelli per ogni variabile di interesse potrebbe risultare potenzialmente troppo dispendioso. In generale, se le osservazioni sono correlate e la media del processo non cambia troppo velocemente di livello, la carta EWMA con un valore opportuno di λ costituisce un eccellente previsore. Un metodo che si è rilevato utile nel trattare dati autocorrelati, è quello di descrivere direttamente la struttura di correlazione con appropriati modelli di serie temporali, con i quali rimuovere la presenza di autocorrelazione e in seguito applicare una carta di controllo (EWMA) ai residui. Quest'ultimi, rappresentati dalla formula

$$e_t = y_t - \hat{y}_t$$

sono indipendenti e hanno distribuzione Normale con media nulla e deviazione standard σ costante. Si può dunque procedere alla realizzazione della carta di controllo applicata ai residui i quali dovrebbero essere indipendenti e identicamente distribuiti, in assenza di cause speciali, mentre dovrebbero fornirne una certa evidenza in presenza di qualche causa speciale.

Infatti l'approccio che è stato seguito per questo lavoro di tesi, è stato:

1. Descrivere la struttura di correlazione con un appropriato modello di serie temporali (si sono trattati in modo particolare, modelli Autoregressivi AR);
2. Rimuovere la presenza di autocorrelazione tramite il modello individuato, tramite il calcolo dei residui;
3. Applicare la carta di controllo EWMA direttamente sui residui (sotto l'assunzione che il modello stimato sia perfetto).

Siccome il modello deve essere stimato dai dati, gli errori di stima saranno inevitabili. Al fine di “limitare” questi errori di stima, nel capitolo successivo saranno presentati alcuni metodi di correzione di stima dei parametri e ne saranno illustrati gli effetti sulle prestazioni della carta di controllo EWMA.

Capitolo 3

Tecniche per correggere gli stimatori autoregressivi

3.1 Proprietà degli stimatori di un processo ARMA

In generale, un processo ARMA opportunamente identificato dalle osservazioni deve essere stimato nei parametri e sottoposto a verifica perché sia utilizzabile per uno scopo definito. Come già indicato nel paragrafo 1.6 del primo capitolo, gli stimatori desiderabili si ottengono con il metodo della massima verosimiglianza (definiti stimatori ML). Diversi studi, in particolare Whittle (1953), hanno dimostrato che l'intera teoria asintotica degli stimatori ML per una numerosità sufficientemente elevata, si applica anche nel caso delle serie temporali. In particolare gli stimatori risultano essere consistenti, efficienti e sufficienti per distribuzioni Gaussiane e risultano asintoticamente Normali ed efficienti anche per famiglie di ripartizione non Gaussiane. Per quanto efficienti possano essere gli algoritmi utilizzati nella fase di stima, occorre cautelare circa il loro utilizzo. Si pensi al fatto che si possono riscontrare valori anomali isolati, che possono creare distribuzioni degeneri nella funzione di verosimiglianza. In questo caso il rimedio più semplice consiste nell'individuare ed eliminare tali dati, mediando le osservazioni o introducendo variabili artificiali per una loro

spiegazione del modello. Oppure l'algoritmo si ferma essendo giunto a un punto di massimo relativo che potrebbe corrispondere a valori ancora distanti dal massimo assoluto. Infine l'algoritmo potrebbe procedere all'infinito trovandosi di fronte a parecchi valori del parametro per i quali la verosimiglianza è costante, e si ferma solo se è previsto un numero massimo di iterazioni.

Tralasciando queste problematiche in altre sedi, in presenza di un processo ARMA stazionario invertibile e causale, le stime di massima verosimiglianza sono equivalenti, asintoticamente, a quelle ottenute con il metodo dei minimi quadrati. Tuttavia per campioni finiti, si prediligono gli stimatori ML, risultando più efficienti. Una proprietà constatata di questi stimatori, è la loro distribuzione asintotica alla Normale. Ad esempio per un modello autoregressivo di ordine 1 a media nulla, in breve AR(1), lo stimatore ML del parametro risulterà:

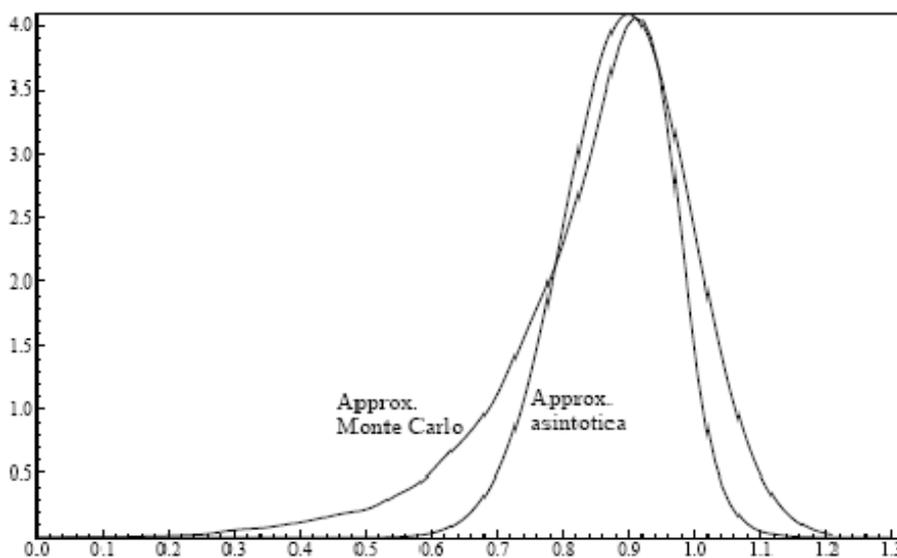
$$\hat{\phi}_n \approx N(\phi, (1 - \phi^2)/n)$$

Purtroppo le proprietà per campioni finiti degli stimatori dei parametri dei processi ARMA sono virtualmente impossibili da derivare analiticamente. Però è noto che le stime dei parametri autoregressivi sono *distorte* in maniera tale che le radici inverse dell'equazione caratteristica tendono a essere mediamente più piccole delle radici vere. Quindi la distorsione cresce con il crescere delle radici inverse, e può essere notevole quando i valori delle radici sono prossimi all'unità. Per esempio, sempre per un AR(1) con media nulla e coefficiente autoregressivo $\phi = 0.9$, la distorsione è circa -0.07 per un campione di 20 osservazioni e scende a circa -0.02 per un campione di 100 osservazioni. La distorsione aumenta notevolmente quando si stima anche la costante: -0.2 per un campione di 20 osservazioni e -0.04 per un campione di 100 osservazioni. Quando $\phi = 0.99$ e la costante viene stimata, si ha una distorsione di -0.23 per $n = 20$ e -0.05 per $n = 100$. Per quanto riguarda la forma della distribuzione della

stima per campioni finiti, si guardi la figura 3.1; è evidente l'asimmetria negativa (positiva quando $\phi < 0$).

Quindi l'uso di stimatori corretti è motivato dal fatto che gli usuali stimatori per un modello AR sono non corretti per campioni poco numerosi. Studi passati hanno dimostrato che stimatori distorti possono procurare effetti negativi sull'accuratezza degli intervalli di previsione, soprattutto nei piccoli campioni. Però non sempre l'applicazione di uno stimatore corretto porta ad un miglioramento perché la correzione potrebbe aumentare la variabilità della previsione abbastanza da "falsare" la riduzione della distorsione, quindi incrementando l'errore quadratico medio di previsione. Inoltre la buona riuscita di questa tecnica dipende anche da quale tipo di stimatore si sta applicando. Quello che si vorrà dimostrare successivamente è vedere l'effetto di questa correzione sui parametri del modello esaminato e di conseguenza, anche sulle applicazioni di qualità.

Figura 3.1 – Approssimazione asintotica e Monte Carlo della distribuzione di $\hat{\phi}_n$ per $n = 20$



3.2 Tecniche per la correzione

In questo paragrafo verranno presentati ed analizzati alcune tecniche per la correzione degli stimatori dei parametri di un processo Autoregressivo. Ma prima è debita una premessa. Lo stimatore ML considerato in precedenza, non è l'unico che si può utilizzare per prevedere il vero valore del parametro autoregressivo del modello. E' stato dimostrato che questo stimatore risulta essere il meno distorto fra quelli a disposizione, quindi è sembrato ragionevole porre il confronto con questa tipologia di stimatore.

Tra questi metodi, ritroviamo la tecnica *bootstrap* che può essere utilizzata per ottenere uno stimatore corretto per i parametri un modello AR. Infatti Kilian (1998) propose uno stimatore corretto bootstrap basato sulla media per i parametri autoregressivi. Questo procedimento fornisce uno stimatore asintoticamente valido nel caso in cui il modello AR risulti stazionario. Mentre per i modelli AR a radice unitaria, le proprietà teoriche dello stimatore bootstrap corretto non sono conosciute. Successivamente i ricercatori Roy e Fuller (2001) suggerirono approssimativamente uno stimatore non distorto per i parametri di un modello autoregressivo basato sulla variante delle stime ottenute con il metodo dei minimi quadrati (Least Squares, LS). Questa modifica utilizza la formula asintotica della distorsione per le stime LS del parametro di correlazione, ricavata appunto da Roy e Fuller. Questo stimatore proposto è pressappoco corretto perfino in presenza di radice unitaria. Un'altra tecnica per la correzione della stima fu suggerita da Andrews e Chen (1994). Costoro avanzarono uno stimatore corretto basato sulla mediana per i parametri di un modello AR di ordine p . Questo fu ottenuto usando la funzione mediana per gli stimatori LS, approssimato dalle simulazioni Monte Carlo. Ora saranno analizzati più nel dettaglio, riprendendo parte della teoria di base sui modelli autoregressivi.

Nel caso più generale, un modello autoregressivo di ordine p , $AR(p)$, lo si può considerare in questa forma:

$$Y_t = \mu' + \beta' + v_t$$

$$v_t = \phi_1 v_{t-1} + \dots + \phi_p v_{t-p} + \varepsilon_t$$

dove $v_t \approx iid(0, \sigma^2)$.

Il modello può essere riscritto come segue:

$$Y_t = \mu + \beta + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t \quad \text{con}$$

$$\mu = \mu'(1 - \phi_1 - \dots - \phi_p) + (\phi_1 + 2\phi_2 + \dots + p\phi_p)\beta'$$

$$\beta = \beta'(1 - \phi_1 - \dots - \phi_p)$$

Di conseguenza si otterrà:

$$Y_t = \mu + \beta t + \phi_1 Y_{t-1} + \iota_1 \Delta Y_{t-1} + \dots + \iota_{p-1} \Delta Y_{t-p+1} + \varepsilon_t$$

con $\Delta = 1-B$ di cui B è l'operatore di ritardo e nel caso generale, $\beta = 0$. Si noti che i parametri delle equazioni precedenti sono collegati così

$$\phi = \phi_1 + \dots + \phi_p \quad \text{e}$$

$$\phi_1 = \phi + \iota_1$$

$$\phi_j = -\iota_{j-1} + \iota_j \quad \text{per } 2 \leq j \leq p-1$$

$$\phi_j = -\iota_{j-1} \quad \text{per } j=p$$

Il parametro ϕ è il parametro di correlazione e misura, appunto, il grado di correlazione del modello AR(p); se questo è uguale a 1, il modello ha radice unitaria (vd. Capitolo 1, par. 4). Date le osservazioni della serie temporale, gli stimatori ricavati con il metodo dei minimi quadrati e i residui associati sono rispettivamente

$$\hat{\gamma} = (\hat{\mu}, \hat{\beta}, \hat{\phi}_1, \dots, \hat{\phi}_p) \text{ e } \{\hat{\varepsilon}_t\}_{t=p+1}^n$$

Riprendendo la procedura proposta da Kilian, in un contesto di modelli autoregressivi, la tecnica bootstrap può essere eseguita dal ricampionamento dei residui e dalla creazione di un numero elevato di pseudo dati utilizzando i coefficienti stimati e i residui ricampionati. Così è replicata la struttura di dipendenza, presente nella serie temporale osservata. Nello specifico vengono generati pseudo dati basati sul modello AR ritardato come:

$$Y_t^* = \hat{\mu} + \hat{\beta}t + \hat{\phi}_1 Y_{t+1}^* + \dots + \hat{\phi}_p Y_{t+p}^* + \varepsilon_t^*$$

dove ε_t^* è un'estrazione casuale con reinserimento dai residui stimati. L'uso di un modello ritardato è utilizzato per includere la condizionalità delle previsioni AR all'interno delle stime bootstrap distorte. La procedura precedente può essere ripetuta più volte affinché N serie sono state generate dagli pseudo dati, dalle quali si possono ottenere le stime bootstrap per γ , denominate

$$\{\hat{\gamma}^*(j)\}_{j=1}^N.$$

La distorsione di $\hat{\gamma}$ si può stimare così:

$$\text{Bias}(\hat{\gamma}) = \bar{\gamma}^* - \hat{\gamma}$$

dove $\bar{\gamma}^* = \text{mean}\left(\{\hat{\gamma}^*(j)\}_{j=1}^N\right)$

Quindi le stime corrette per γ risultano

$$\hat{\gamma}^c = \hat{\gamma} - Bias(\hat{\gamma})$$

Però questo tipo di correzione può portare gli stimatori verso la non stazionarietà dello spazio parametrico. Per ovviare a questo inconveniente, Kilian propose un aggiustamento che riporti le stime ad essere stazionarie. Se lo stimatore non distorto calcolato racchiude della non stazionarietà, allora:

$$\begin{aligned} \delta_1 &= 1 \quad , \quad \Delta_1 = Bias(\hat{\gamma}) \quad e \\ \hat{\gamma}_i^c &= \hat{\gamma} - \Delta_i. \\ \text{Pongo } \Delta_{i+1} &= \delta_i \Delta_i \quad , \\ \delta_{i+1} &= \delta_i - 0.01 \quad \text{per } i=1,2,3,.. \end{aligned}$$

Dopodiché viene ripetuta questa procedura finché gli stimatori non distorti soddisfano la condizione di stazionarietà. Ad esempio per un modello AR(1), se si ottiene $\hat{\phi}_1 = 0.95$ e $\hat{\phi}_1^c = 1.05$, quest'ultima sarà "aggiustata" a 0,99.

Invece lo stimatore Roy-Fuller fornisce una semplice modifica delle stime ottenute con il metodo dei minimi quadrati. Questo procedimento pone:

$$\begin{aligned} \tilde{\phi} &= \min(\ddot{\phi}, 1) \quad \text{dove} \\ \ddot{\phi} &= \dot{\phi} + [C_p(\hat{t}_{-1})] \hat{\sigma}_1 \end{aligned}$$

In questo caso $\dot{\phi}$ è lo stimatore LS per il coefficiente di \hat{y}_{t-1} nella regressione di \hat{y}_t su \hat{y}_{t-1} , $\Delta\hat{y}_{t-1}$, ..., $\Delta\hat{y}_{t-p+1}$, dove \hat{y}_t sono i residui LS della regressione di Y_t su (1, t) e $\hat{\sigma}_1^{-1}$ è lo standard error di $\hat{\phi}$. Inoltre si noti che il test statistico della radice unitaria è definito come

$$\begin{aligned} \hat{t}_1 &= (\dot{\phi} - 1) \hat{\sigma}_1^{-1} \quad e \\ \hat{t}_{-1} &\text{ è il test per l'ipotesi } \phi = -1. \end{aligned}$$

Intanto le due funzioni $C_p(\hat{\tau}_1)$ e $C_{-p}(\hat{\tau}_{-1})$ controllano la direzione in cui la correzione è condotta e ne determinano la grandezza. Queste sono costruite in modo tale che $\tilde{\phi}$ risulti in modo approssimato una mediana non distorta, rispettivamente per $\phi = 1$ e -1 . Inoltre Roy e Fuller proposero per la funzione $C_p(\hat{\tau}_1)$ la seguente forma, che è impiegata per la formula asintotica della distorsione per $\hat{\phi}$:

$$\begin{aligned} [C_p(\hat{\tau}_1)] &= -\tau_{\text{MED}} + d_n(\hat{\tau}_1 - \tau_{\text{MED}}) && \text{se } \hat{\tau}_1 > \tau_{\text{MED}} \\ &= I_p(n^{-1}\hat{\tau}_1) - 3[\hat{\tau}_1 + k(\hat{\tau}_1 + K)]^{-1} && \text{se } -K < \hat{\tau}_1 \leq \tau_{\text{MED}} \\ &= I_p(n^{-1}\hat{\tau}_1) - 3[\hat{\tau}_1]^{-1} && \text{se } -\sqrt{3n} < \hat{\tau}_1 \leq -K \\ &= 0 && \text{se } \hat{\tau}_1 \leq -\sqrt{3n} \end{aligned}$$

, dove I_p è la parte intera di $0.5(p + 1)$, τ_{MED} è la mediana della distribuzione limite di $\hat{\tau}_1$ quando $\phi = 1$, e $k = [3n - \tau_{\text{MED}}^2(I_p + n)] \cdot [\tau_{\text{MED}}(K + \tau_{\text{MED}})(I_p + n)]^{-1}$, mentre il valore di K e d_n è preso tra 5 e 0.29 come proposto nel lavoro di Roy e Fuller. Una volta che è stato trovato il valore di $\tilde{\phi}$, $(\mu, \beta, \psi_1, \dots, \psi_{p-1})$ possono essere stimate dalla regressione di $Y_t - \tilde{\phi}Y_{t-1}$ su $(1, t, \Delta Y_{t-1}, \dots, \Delta Y_{t-p+1})$. Si noti che β assume valore zero nel momento in cui $\tilde{\phi} = 1$. Quindi utilizzando le relazioni precedentemente citate per il modello autoregressivo, gli stimatori corretti basati sul metodo proposto da Roy e Fuller possono essere “trasformati” in modo appropriato e vengono indicati con $\tilde{\gamma} = (\tilde{\mu}, \tilde{\beta}, \tilde{\phi}_1, \dots, \tilde{\phi}_p)$.

Infine tra gli stimatori proposti, è stato menzionato anche quello di Andrews e Chen, che cerca di trovare in via approssimata lo stimatore corretto basato sulla mediana per il parametro di correlazione del modello. Dato $m(\phi)$ come funzione mediana di $\hat{\phi}$, che è strettamente crescente sullo spazio parametrico di

$\phi \in (-1,1)$. Allora approssimativamente lo stimatore non distorto basato sulla mediana per $\hat{\phi}_U$ sarà dato da:

$$\begin{aligned}\hat{\phi}_U &= 1 && \text{se } \hat{\phi} > m(1) \\ &= m^{-1}(\hat{\phi}) && \text{se } m(-1) < \hat{\phi} \leq m(1) \\ &= -1 && \text{se } \hat{\phi} \leq m(-1)\end{aligned}$$

, dove

$$m(-1) = \lim_{\phi \rightarrow -1} m(\phi)$$

E m^{-1} è la funzione inversa di $m(\phi)$. Questa funzione mediana si può calcolarla utilizzando il metodo di simulazione Monte Carlo, come descritto nel lavoro di Andrews e Chen. Come il metodo di simulazione può determinare la funzione $m(\phi)$ su una griglia di valori per ϕ sull'intervallo -1 e 1, una lineare interpolazione è condotta al fine di approssimare il valore di $m^{-1}(\hat{\phi})$, come descritto da Andrews (1993). Cosa importante che va sottolineata è che questo stimatore risulta esattamente la mediana corretta per un modello AR(1), mentre è approssimato se l'ordine del modello autoregressivo è maggiore di 1. Una volta che recuperato il valore di $\hat{\phi}_U$, $(\mu, \beta, \psi_1, \dots, \psi_{p-1})$ possono essere stimate dalla regressione di $Y_t - \hat{\phi}_U Y_{t-1}$ su $(1, t, \Delta Y_{t-1}, \dots, \Delta Y_{t-p+1})$. Come nel metodo di Roy e Fuller, β diventa uguale a zero nel momento in cui $\hat{\phi}_U = 1$. Di conseguenza la procedura appena descritta può essere iterata fino a convergere, come descritto nel lavoro sempre di Andrew e Chen (1994). Infine per ottenere le stime non distorte, si lavora come nel caso degli stimatori di Roy e Fuller, e vengono indicate $\hat{\gamma}_U = (\hat{\mu}_U, \hat{\beta}_U, \hat{\phi}_{1,U}, \dots, \hat{\phi}_{p,U})$. Dopo una esauriente presentazione di alcuni stimatori corretti, il nostro compito sarà quello di vedere la loro applicazione e di analizzare l'effetto che scaturiscano in applicazioni di qualità.

Capitolo 4

Simulazione e risultati

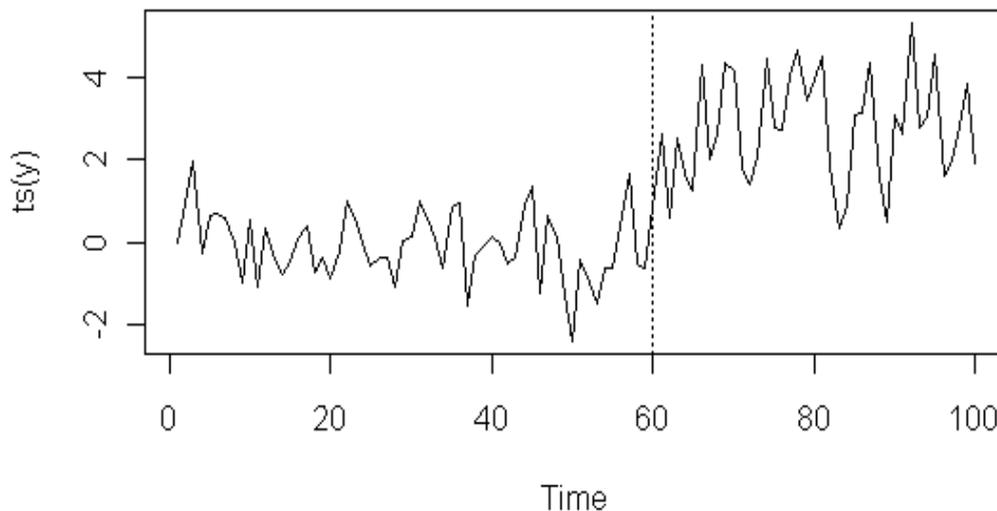
4.1 Scelta dei parametri

Dopo aver appreso ed analizzato le caratteristiche di alcuni stimatori corretti per i parametri di un processo ARMA, è giunto il momento di valutarli quando utilizzati in applicazioni di qualità. L'obiettivo è quello di esaminare l'effetto della correzione degli stimatori che caratterizzano il modello preso in esame per la costruzione di carte di controllo EWMA e vedere se questa correzione, porta a risultati soddisfacenti. Innanzitutto bisogna definire i parametri che verranno utilizzati nella simulazione, sia per la serie temporale sia per la costruzione della carta di controllo. Come descritto nei capitoli precedenti, il numero delle variabili che interagiscono è abbastanza elevato, come di conseguenza il numero di possibilità che avremmo in sede di simulazione. Quindi si sono scelti determinati valori per alcuni parametri, definendo dei casi limite che andranno a rappresentare l'intero "pacchetto" di opportunità.

Per l'analisi della serie temporale, è stato utilizzato un modello di tipo autoregressivo (AR) di ordine 1 con $\mu=0$ e $\sigma=1$. Questo modello è caratterizzato da un solo parametro (φ) che può variare tra i valori -1 e 1, ma nelle simulazioni realizzate sono stati considerati solo 3 casi limite: forte correlazione negativa (-

0,75), forte correlazione positiva (0,75) e debole correlazione positiva (0,3). Per ognuno dei modelli descritti, sono state indagate dimensioni campionarie differenti: 50, 100 e 200 unità. Parte di queste osservazioni da un certo punto in poi, subiscono uno slittamento (definito “shift”), portando ad un cambiamento della media del processo. Come ben visibile nell’esempio della figura 4.1, la prima parte della serie si definisce *in controllo*, mentre la seconda parte, che ha subito l’effetto dello shift, *fuori controllo*. Si noti che il numero di osservazioni mandate fuori controllo è stato scelto in modo arbitrario ma di ugual identità per tutti i casi trattati. A riguardo l’obiettivo primario è stato quello di aver un certo numero di dati fuori controllo tale che, su questi, potesse essere evidente il comportamento degli stimatori utilizzati.

Figura 4.1 – Salto di media all’interno di un modello AR(1)



Per lo shift si potevano considerare infiniti valori ma anche in questa circostanza, si sono definite due ipotesi: shift = 1 e shift = 3. Nel primo caso si è

scelto un salto della media piccolo, dove quasi sicuramente la carta di controllo EWMA segnalerà questa situazione con un certo ritardo. Nel secondo caso il salto della media pare più sostanzioso e sarà evidenziato con tempestività dalla carta di controllo. Anche per le variabili che riguardano la costruzione della carta EWMA, sono stati considerati dei casi limite: per il parametro λ , costante che varia tra 0 e 1 e che definisce quanto peso assegnare all'ultima osservazione, vengono considerati i valori 0,1 (basso) e 0,4 (medio - alto). Infine per ogni simulazione sono stati analizzati dei fattori che determinano l'efficienza della carta di controllo EWMA. Come già descritti dettagliatamente nel Capitolo 2, questi sono: l'*Average Run Length* (ARL), la proporzione dei *fuori controllo non segnalati* (FCNS) e la proporzione dei *falsi allarmi* (FALL). Quindi combinando i valori assegnati alle variabili con gli stimatori corretti proposti al capitolo precedente, sono state ricavate 12 tabelle che andremo ad analizzare e commentare.

4.2 Interpretazione e discussione dei risultati

I risultati ottenuti sono sostanzialmente concordi e delineano in modo univoco il comportamento degli stimatori corretti, applicati nei modelli autoregressivi considerati. Per giungere a queste conclusioni, è stato necessario l'ausilio del software statistico *R* (versione 2.8.0), con il quale si sono svolte tutte le procedure. Per realizzare queste elaborazioni, si è voluto seguire un "ordine sistematico" che alla fine portasse alla costruzione della carta di controllo EWMA e dei fattori che ne quantificassero l'efficienza in termini di qualità. In particolare:

1. Innanzitutto sono state definite le variabili riguardanti il campione (numerosità campionaria, osservazioni in controllo e fuori controllo, numero di simulazioni), quelle che caratterizzano il modello autoregressivo (il parametro ϕ , lo shift δ), quelle pertinenti la carta di controllo EWMA (λ e h per i limiti di controllo e una statistica w basata sui residui per la costruzione della carta) ed infine quelle che sono servite per confrontare gli stimatori proposti (i vettori $\hat{\mu}$, $\hat{\sigma}$, $\hat{\phi}$ e $Var(\hat{\phi})$ per il modello autoregressivo e i vettori ARL, FCNS, FALL per la carta di controllo EWMA).
2. Successivamente sono state create due funzioni: la prima (denominata *ARsim*) che permettesse di simulare dati da un modello AR(1), la seconda (detta *ARres*) che calcolasse i residui proprio dal modello precedente.
3. All'interno di un ciclo di simulazioni (nel nostro caso 1000), è stata richiamata la funzione *ARsim* per la simulazione dei dati, che sono stati suddivisi parte in controllo e parte in fuori controllo (ai quali è stato aggiunto lo shift). Sulla metà di quelli in controllo, è stato stimato il modello nella maniera usuale, ovvero con gli stimatori di massima verosimiglianza. La scelta di aver utilizzato solo la metà dei dati in controllo per la stima del modello è stata ponderata al fine di esaminare

come si comportavano le stime nella seconda parte di osservazioni in controllo, oltre a quelle fuori controllo. Una volta raccolte le stime del modello, proprio su queste si sono calcolati i residui tramite la funzione ARres e in seguito la statistica w basata, appunto, sui residui. Quest'ultima ha forma:

$$w_t = \lambda \cdot e_{t-1} + (1 - \lambda) \cdot w_{t-1}$$

con $w_0 = 0$

dove e_{t-1} rappresenta i residui al tempo $t-1$. Infine è stato ripetuto lo stesso procedimento ma utilizzando la procedura per il computo delle stime corrette.

4. Dopo il ciclo di simulazioni, sia per le stime usuali che per le stime corrette, sono stati definiti i limiti di controllo UCL e LCL basati sulle stime e lo *shift standardizzato*, inteso come:

$$\hat{\delta} = \delta \cdot \sqrt{1 - (1 - \hat{\phi})^2}.$$

5. In conclusione è stato creato un altro ciclo di simulazioni dove, sempre per entrambe le stime, si sono conteggiati ARL, i fuori controllo non segnalati e i falsi allarmi.

Ovviamente alla fine di questo elenco di procedure, si sono eseguite analisi grafiche per identificare meglio l'effetto degli stimatori, ma non le andremo ad elencare. Per i dettagli rimando all'Appendice dove è stato inserito il *codice R*, con il quale si sono ottenuti i risultati. Prima di discutere i risultati, riassunti nelle tabelle che verranno, è doverosa una premessa. Come si potrà ravvisare, si

farà riferimento ad uno solo degli stimatori proposti nel capitolo precedente: lo stimatore corretto bootstrap basato sulla media (d'ora in poi $\tilde{\phi}_{Boot}$). Questo perché l'effetto che producono gli stimatori corretti nelle applicazioni di qualità analizzate, è sostanzialmente identico fra loro. Quindi per semplicità di simulazione e di interpretazione dei risultati, si è deciso di utilizzarne uno solo in rappresentanza degli altri. La scelta è ricaduta su $\tilde{\phi}_{Boot}$ perché, come dimostrato nel lavoro di J. H. Kim (2001), questo stimatore corretto si comporta leggermente meglio degli altri su piccoli campioni, in ambito di intervalli di previsioni per modelli autoregressivi. Come ribadito in precedenza, i risultati delle simulazioni sono state riassunte nelle seguenti tabelle.

Tabella 4.1 – Modello AR(1) con $\varphi = -0.75$, $\lambda = 0.1$ e shift = 1

	Unità campione	\bar{x}	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\phi}$	$Var(\hat{\phi})$	ARL	FCNS	FALL
$\hat{\phi}_{ML}$	50(30)	-0,078	0,001	0,886	-0,64	0,039	6,624	15,479	0,270
	100(60)	0,035	0,002	0,947	-0,69	0,018	25,004	38,261	0,022
	200(120)	-0,001	-0,002	0,974	-0,72	0,007	70,832	79,672	0,001
$\tilde{\phi}_{Boot}$	50(30)	-0,078	0,002	0,956	-0,72	0,058	9,238	15,453	0,436
	100(60)	0,035	0,001	0,982	-0,75	0,023	26,052	37,380	0,088
	200(120)	-0,001	-0,003	0,991	-0,75	0,008	66,714	79,235	0,010

Tabella 4.2 – Modello AR(1) con $\varphi = -0.75$, $\lambda = 0.1$ e shift = 3

	Unità campione	\bar{x}	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\phi}$	$Var(\hat{\phi})$	ARL	FCNS	FALL
$\hat{\phi}_{ML}$	50(30)	-0,002	-0,002	0,885	-0,63	0,041	1,322	1,097	0,207
	100(60)	-0,012	-0,003	0,943	-0,69	0,017	1,812	5,183	0,027
	200(120)	0,020	-0,004	0,971	-0,72	0,008	2,512	25,102	0,000
$\tilde{\phi}_{Boot}$	50(30)	-0,002	-0,004	0,950	-0,71	0,062	1,738	2,352	0,451
	100(60)	-0,012	-0,004	0,970	-0,75	0,021	2,235	7,519	0,081
	200(120)	0,020	-0,004	0,984	-0,75	0,010	3,625	27,881	0,002

Tabella 4.3 – Modello AR(1) con $\varphi = -0.75$, $\lambda = 0.4$ e shift = 1

	Unità campione	\bar{x}	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\phi}$	$Var(\hat{\phi})$	ARL	FCNS	FALL
$\hat{\phi}_{ML}$	50(30)	0,137	0,001	0,890	-0,64	0,038	3,191	11,313	1,168
	100(60)	0,099	0,002	0,947	-0,69	0,018	5,340	29,792	0,584
	200(120)	-0,024	-0,001	0,972	-0,72	0,008	10,140	69,228	0,241
$\tilde{\phi}_{Boot}$	50(30)	0,137	0,001	0,963	-0,73	0,057	5,119	12,140	1,196
	100(60)	0,099	0,003	0,981	-0,74	0,023	8,662	29,904	0,735
	200(120)	-0,024	-0,001	0,987	-0,75	0,009	13,940	68,243	0,357

Tabella 4.4 – Modello AR(1) con $\varphi = -0.75$, $\lambda = 0.4$ e shift = 3

	Unità campione	\bar{x}	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\phi}$	$Var(\hat{\phi})$	ARL	FCNS	FALL
$\hat{\phi}_{ML}$	50(30)	0,068	-0,003	0,896	-0,65	0,038	1,144	0,370	1,134
	100(60)	-0,002	-0,001	0,955	-0,69	0,018	1,305	0,808	0,521
	200(120)	-0,006	-0,004	0,970	-0,72	0,007	1,438	1,764	0,290
$\tilde{\phi}_{Boot}$	50(30)	0,068	-0,003	0,976	-0,74	0,057	1,349	0,886	1,135
	100(60)	-0,002	-0,002	0,996	-0,74	0,023	1,406	1,491	0,613
	200(120)	-0,006	-0,005	0,982	-0,75	0,009	1,488	2,374	0,422

Tabella 4.5 – Modello AR(1) con $\varphi = 0.3$, $\lambda = 0.1$ e shift = 1

	Unità campione	\bar{x}	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\phi}$	$Var(\hat{\phi})$	ARL	FCNS	FALL
$\hat{\phi}_{ML}$	50(30)	-0,065	0,001	0,883	0,15	0,062	12,625	18,187	0,332
	100(60)	-0,149	-0,001	0,949	0,22	0,030	36,740	39,700	0,019
	200(120)	-0,003	0,008	0,972	0,27	0,016	80,227	79,979	0,002
$\tilde{\phi}_{Boot}$	50(30)	-0,065	0,001	0,941	0,27	0,102	13,874	18,044	0,610
	100(60)	-0,149	-0,002	0,985	0,28	0,039	35,623	39,425	0,084
	200(120)	-0,003	0,008	0,986	0,30	0,018	78,502	79,901	0,007

Tabella 4.6 – Modello AR(1) con $\varphi = 0.3$, $\lambda = 0.1$ e shift = 3

	Unità campione	\bar{x}	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\phi}$	$Var(\hat{\phi})$	ARL	FCNS	FALL
$\hat{\phi}_{ML}$	50(30)	-0,074	0,018	0,885	0,16	0,061	2,665	10,116	0,342
	100(60)	-0,118	-0,018	0,943	0,23	0,030	12,279	33,653	0,025
	200(120)	0,023	0,005	0,975	0,26	0,015	52,550	78,315	0,002
$\tilde{\phi}_{Boot}$	50(30)	-0,074	0,014	0,950	0,28	0,100	6,188	12,760	0,583
	100(60)	-0,118	-0,019	0,974	0,30	0,040	18,630	34,098	0,100
	200(120)	0,023	0,005	0,993	0,30	0,018	56,135	77,640	0,008

Tabella 4.7 – Modello AR(1) con $\varphi = 0.3$, $\lambda = 0.4$ e shift = 1

	Unità campione	\bar{x}	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\phi}$	$Var(\hat{\phi})$	ARL	FCNS	FALL
$\hat{\phi}_{ML}$	50(30)	0,041	0,013	0,874	0,15	0,061	6,453	15,526	1,257
	100(60)	0,053	-0,005	0,940	0,22	0,031	17,250	36,991	0,586
	200(120)	0,034	0,003	0,974	0,27	0,015	42,672	78,170	0,266
$\tilde{\phi}_{Boot}$	50(30)	0,047	0,005	0,921	0,27	0,100	8,734	16,028	1,570
	100(60)	0,053	-0,002	0,968	0,29	0,040	20,437	36,870	0,785
	200(120)	0,034	0,003	0,991	0,30	0,017	47,635	77,899	0,388

Tabella 4.8 – Modello AR(1) con $\varphi = 0.3$, $\lambda = 0.4$ e shift = 3

	Unità campione	\bar{x}	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\phi}$	$Var(\hat{\phi})$	ARL	FCNS	FALL
$\hat{\phi}_{ML}$	50(30)	-0,020	0,010	0,876	0,14	0,064	1,293	5,561	1,331
	100(60)	-0,056	0,001	0,944	0,22	0,029	1,847	20,176	0,592
	200(120)	0,118	0,014	0,969	0,26	0,016	3,598	56,998	0,287
$\tilde{\phi}_{Boot}$	50(30)	-0,020	0,008	0,923	0,26	0,106	2,539	8,879	1,568
	100(60)	-0,056	0,003	0,976	0,29	0,037	4,429	24,746	0,732
	200(120)	0,118	0,013	0,981	0,30	0,018	6,499	59,360	0,440

Tabella 4.9 – Modello AR(1) con $\varphi = 0.75$, $\lambda = 0.1$ e shift = 1

	Unità campione	\bar{x}	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\phi}$	$Var(\hat{\phi})$	ARL	FCNS	FALL
$\hat{\phi}_{ML}$	50(30)	-0,228	0,033	0,889	0,47	0,061	15,005	18,584	0,412
	100(60)	-0,101	0,030	0,940	0,60	0,025	38,525	39,790	0,028
	200(120)	0,296	0,013	0,971	0,68	0,012	80,618	79,983	0,000
$\tilde{\phi}_{Boot}$	50(30)	-0,228	0,041	0,952	0,65	0,090	16,043	18,310	1,209
	100(60)	-0,101	0,026	0,974	0,72	0,033	37,674	39,473	0,181
	200(120)	0,296	0,020	0,989	0,74	0,014	79,779	79,957	0,004

Tabella 4.10 – Modello AR(1) con $\varphi = 0.75$, $\lambda = 0.1$ e shift = 3

	Unità campione	\bar{x}	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\phi}$	$Var(\hat{\phi})$	ARL	FCNS	FALL
$\hat{\phi}_{ML}$	50(30)	0,861	0,000	0,887	0,49	0,057	5,397	14,834	0,437
	100(60)	-0,104	-0,025	0,942	0,60	0,024	22,837	38,290	0,033
	200(120)	-0,052	-0,015	0,978	0,68	0,011	73,094	79,778	0,000
$\tilde{\phi}_{Boot}$	50(30)	0,861	0,035	0,949	0,67	0,082	9,320	16,227	1,271
	100(60)	-0,104	-0,038	0,980	0,72	0,033	26,421	38,275	0,172
	200(120)	-0,052	-0,021	1,003	0,74	0,012	72,171	79,585	0,012

Tabella 4.11 – Modello AR(1) con $\varphi = 0.75$, $\lambda = 0.4$ e shift = 1

	Unità campione	\bar{x}	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\phi}$	$Var(\hat{\phi})$	ARL	FCNS	FALL
$\hat{\phi}_{ML}$	50(30)	0,232	0,013	0,872	0,46	0,060	9,063	16,319	1,555
	100(60)	-0,074	-0,019	0,942	0,61	0,024	24,134	38,188	0,708
	200(120)	0,329	-0,003	0,973	0,68	0,009	60,059	79,250	0,291
$\tilde{\phi}_{Boot}$	50(30)	0,232	0,051	0,913	0,64	0,087	11,368	16,668	2,507
	100(60)	-0,074	-0,028	0,982	0,73	0,032	27,072	38,103	1,071
	200(120)	0,329	-0,009	0,992	0,74	0,011	61,395	79,133	0,403

Tabella 4.12 – Modello AR(1) con $\varphi = 0.75$, $\lambda = 0.4$ e shift = 3

	Unità campione	\bar{x}	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\phi}$	$Var(\hat{\phi})$	ARL	FCNS	FALL
$\hat{\phi}_{ML}$	50(30)	-0,573	-0,002	0,878	0,47	0,060	2,058	11,222	1,392
	100(60)	-0,898	-0,019	0,941	0,61	0,026	5,662	32,978	0,662
	200(120)	-0,134	-0,049	0,974	0,68	0,010	19,209	75,922	0,297
$\tilde{\phi}_{Boot}$	50(30)	-0,573	0,008	0,935	0,64	0,090	4,552	14,233	2,170
	100(60)	-0,898	-0,017	0,979	0,73	0,035	11,076	34,902	1,081
	200(120)	-0,134	-0,057	0,994	0,75	0,011	28,599	76,556	0,438

Si osservi che in parentesi, accanto alla numerosità campionaria, è indicato il numero di osservazioni in controllo per ogni specifico caso. In generale, si può notare come le tabelle mostrino una similitudine di interpretazione dei risultati, ad ogni livello di valori assunto dal parametro ϕ . Per una migliore comprensione, la discussione sull'effetto degli stimatori non distorti va tenuta parallelamente su due vie: la prima sul modello autoregressivo; la seconda sui fattori che determinano l'efficienza della carta di controllo EWMA.

Come ci si poteva aspettare, e come intravisto sui lavori citati al capitolo precedente, gli stimatori corretti portano ad un miglioramento della stima del parametro del modello AR. Già a basse numerosità campionarie, le stime non distorte sono sensibilmente più vicine al vero valore del parametro; accade la stessa situazione per la stima di σ . Mentre la varianza dello stimatore proposto, soprattutto quando non si ha la possibilità di reperire un campione più numeroso, risulta più elevata di quella dello stimatore ML. Come si può vedere dall'esempio delle figura 4.2, la maggior precisione dello stimatore corretto paga in termini di variabilità. Questo andrà a condizionare i residui e di conseguenza, i

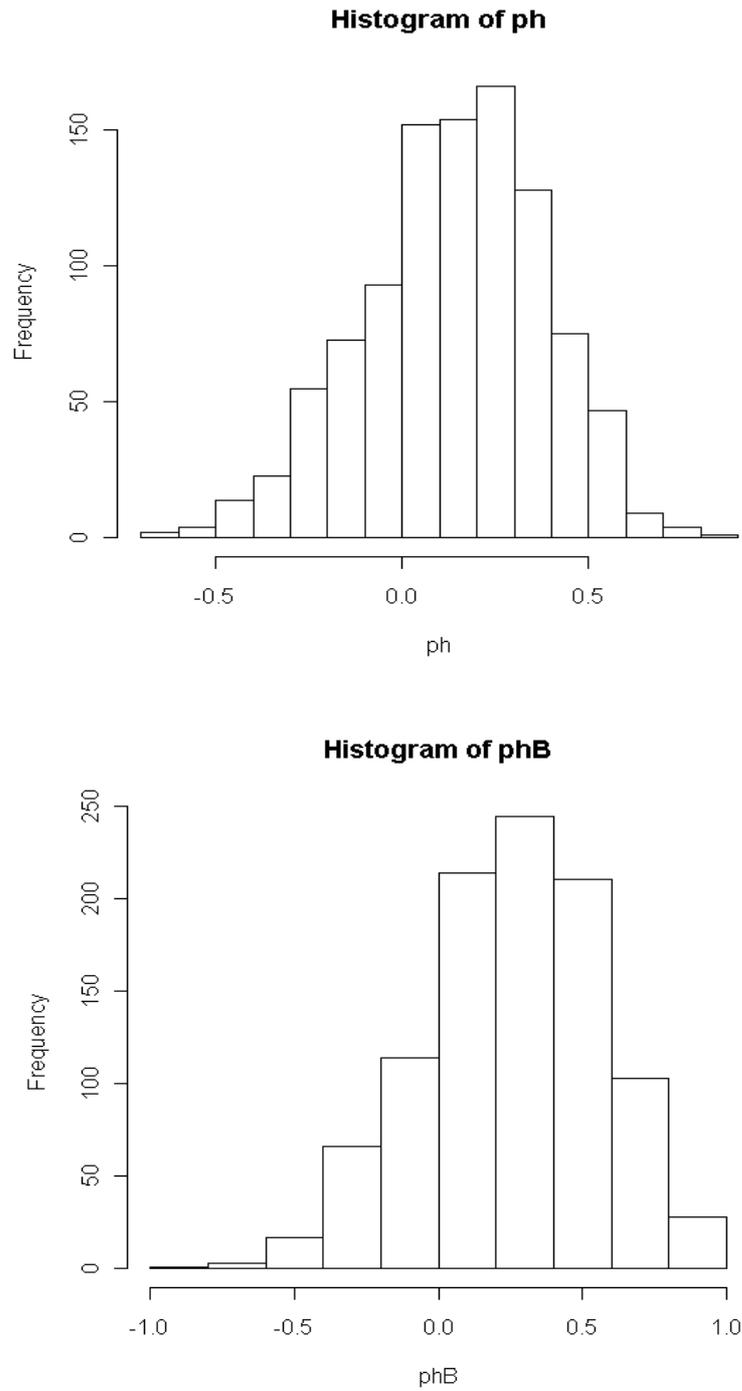
termini che stabiliscono l'efficienza della carta EWMA. Invece per quanto riguarda la media del processo, non sembra esserci grandi differenze tra le stime usuali e quelle proposte.

Nelle applicazioni di qualità, il confronto fra le procedure di stima presentate ha messo in risalto aspetti interessanti. In generale si possono trarre importanti conclusioni dall'analisi delle tabelle, nella fattispecie:

- Al pari della numerosità campionaria, dello shift e del parametro λ all'aumentare del grado di correlazione del modello AR, la carta di controllo EWMA identifica sempre più con un certo ritardo il cambiamento della media, quindi le osservazioni fuori controllo. Infatti passando gradualmente per i 3 valori proposti per il parametro autoregressivo ϕ , per entrambe le procedure di stima, l'ARL tende ad aumentare.
- Si sono ottenuti risultati migliori, in termini di ARL, per valori alti di λ . Precisamente mantenendo inalterato il salto della media, la carta di controllo EWMA sembra reagire meglio per valori di λ pari a 0,4 per entrambe le procedure di stima utilizzate. Questa conseguenza poteva essere attesa dato che il parametro λ rappresenta quanto peso assegnare all'osservazione precedente e quindi, trattando modello autoregressivi, la carta di controllo sarebbe stata più sensibile per valori alti di λ . Inoltre è visibile un leggero miglioramento anche per i fuori controllo non segnalati ma non per i falsi allarmi, che tendono debolmente ad aumentare. Scontato dire che a parità di valore per λ , l'ARL, i FCNS e i FALL peggiorano se lo shift è di minore entità.

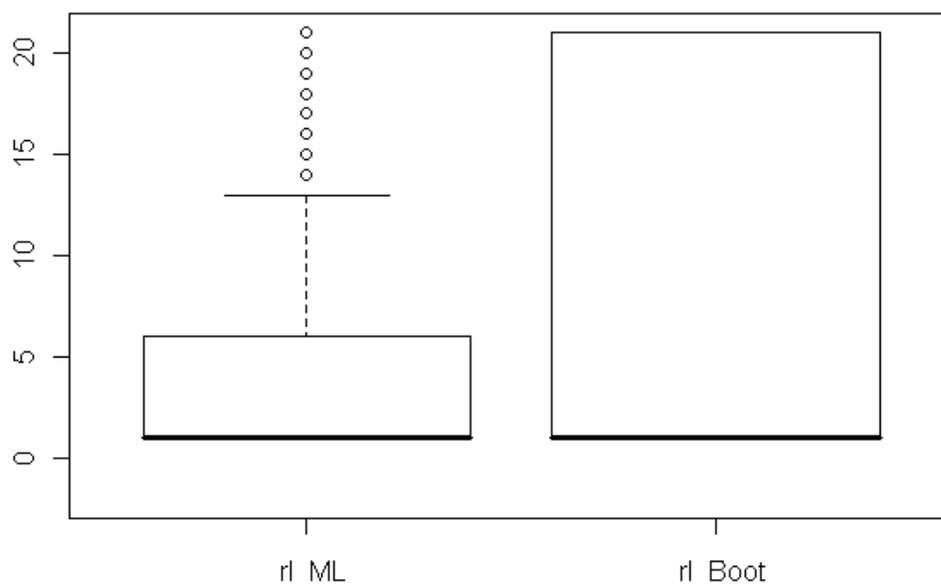
Dopo aver confrontato le tabelle tra loro, e quindi i vari modelli autoregressivi costruiti, vediamo l'effetto degli stimatori avanzati all'interno dei singoli modelli, giungendo all'obiettivo principale del lavoro.

Figura 4.2 – Istogrammi di $Var(\hat{\phi}_{ML})$ e di $Var(\hat{\phi}_{Boot})$ per un modello AR(1) con $\phi = 0.3$, $\lambda = 0.1$ e shift = 3 e numerosità campionaria pari a 50(30).



A parità di numerosità campionaria, si vuole vedere se lo stimatore non distorto si comporta meglio di quello usuale, in applicazioni di qualità. In modo univoco per tutte le tabelle presentate, si può dire che lo stimatore corretto sembrerebbe non portare dei benefici in termini di efficienza alla carta di controllo EWMA. Infatti si può constatare che l'ARL più piccolo, quindi migliore, lo si ottiene quando si utilizza lo stimatore usuale di massima verosimiglianza (esempio figura 4.3). Questa conseguenza sembra essere amplificata maggiormente a basse numerosità campionarie e perfino in presenza di uno shift più alto, ma di minore entità. Per quanto riguarda i fuori controllo non segnalati non ci sono sostanziali differenze se si usano gli stimatori usuali o quelli proposti. Infine per i falsi allarmi sembra esserci un lieve effetto in negativo se si utilizzano le stime corrette. A riguardo, quest'esito sembra ingrandirsi se della numerosità campionaria risulta esigua, come visto per l'ARL.

Figura 4.3 – Boxplot a confronto per l'RL basato su $\hat{\phi}_{ML}$ e per quello basato su $\hat{\phi}_{Boot}$ in un modello AR(1) con $\phi = 0.75$, $\lambda = 0.1$ e shift = 3 e numerosità campionaria pari a 50(30).



Si ritiene che i risultati ottenuti sembrano essere vincolati al numero di unità del campione. Se il campione scelto per la stima dei parametri risulta modesto, le differenze tra i metodi di stima dei parametri autoregressivi tendono ad aumentare a favore dello stimatore di massima verosimiglianza.

Conclusioni

Come accennato nel capitolo precedente, i risultati ottenuti sono notevolmente influenzati dalla variabilità delle stime proposte. Infatti gli stimatori corretti presentano una varianza più elevata degli stimatori usuali e la differenza fra queste due tende ad approssimarsi a zero al crescere delle unità campionarie. Quindi i benefici di non distorsione che portano gli stimatori indicati per i parametri autoregressivi, vengono parzialmente annullati dalla loro elevata varianza. Si può concludere che nonostante la tecnica di correzione della distorsione porti ad una stima più accurata dei parametri del modello AR, in termini di efficienza della carta di controllo EWMA, questo non succede e tendono a preferirsi gli stimatori usuali di massima verosimiglianza. Un'ipotesi futura che si potrebbe suggerire, sarebbe quella di sperimentare nuovi stimatori corretti per i parametri autoregressivi e vedere che effetto hanno sui fattori che determinano l'efficienza di una carta di controllo. In alternativa, si potrebbe pensare di trovare delle metodologie compatibili all'argomento che propongano stimatori non distorti con una variabilità "ridotta", al massimo pari a quella dei consueti stimatori ML.

Appendice

Funzione ARsim per simulare dati da un modello AR

```
ARsim<-function(oldz=0,olda=0,phi,sdinnov,num)
{
  z<-rep(0,num)
  z[1]<-oldz
  for(i in 2:(num))
  {
    newa<-rnorm(n=1,mean=0,sd=sdinnov)
    z[i]<-phi*z[i-1]+newa
    olda<-newa
  }
  return(z)
}
```

, dove:

- *oldz* è il valore iniziale della sequenza simulata
- *olda* è il valore iniziale dell'innovazione
- *phi* è il parametro autoregressivo
- *sdinnov* è la deviazione standard delle innovazioni

- *num* è il numero di dati simulati

Funzione ARres per stimare i residui dal modello precedente

```
ARres<-function(y=dati[1:(length(dati)-1)],ph,mu,sigma,e0=0)
{
  res<-rep(0,length(y))
  res[1]<-e0
  for(i in 2:length(y))
  {
    prev<-ph*(y[i-1]-mu)
    res[i]<-((y[i]-mu)-prev)/sigma
  }
  res<-res[2:length(y)]
  return(res)
}
  olda<-newa
}
  return(z)
}
```

, dove:

- *y* è la serie simulata con la funzione ARsim
- *ph* è il parametro di filtraggio AR
- *mu* è la media della serie simulata *y*
- *sigma* è la deviazione standard delle innovazioni della serie simulata
- *e0* è il valore iniziale che serve per calcolare i residui

Funzione per calcolare il valore critico h

```
h<-xewma.crit(lambda,IC)
```

, dove:

- *lambda* è il parametro che assegna quanto peso attribuire all'ultima osservazione
- *IC* rappresenta il numero di osservazioni in controllo

Ciclo di simulazioni

```
for(i in 1:nsim)
{
  x<-ARsim(0,0,phi,1,ncampione)
  old.seed<-.Random.seed
  xin<-x[1:(IC)]
  xout<-shift+x[(IC+1):ncampione]
  y<-c(xin,xout)
  m<-arima(xin[1:n],c(1,0,0),method="ML")

  ph[i]<-m$coef[1]
  sigma[i]<-sqrt(m$sigma)
  mu[i]<-m$coef[2]

  res<-ARres(y,ph[i],mu[i],sigma[i],0)
  for(t in 2:ncampione)
  {
    w[i,t]<-((lambda*res[t-1])+((1-lambda)*w[t-1]))
  }
}
```

```

B<-BootBC(xin[1:n],1,0,500,type="const",correct="kilian")
.Random.seed<-old.seed

phB[i]<-B$coef[1]
muB[i]<-B$coef[2]/(1-phi)
sigmaB[i]<-sum(B$resid^2)/(n-p-1)

resB<-ARres(y,phB[i],muB[i],sigmaB[i],0)
for(s in 2:ncampione)
{
  wB[i,s]<-((lambda*resB[s-1])+((1-lambda)*wB[s-1]))
}
}

```

, dove:

- *shift* rappresenta il salto della media
- *w* e *wB* sono le statistiche basate sui residui degli rispettivi stimatori sui quali si costruiscono le carte di controllo EWMA
- *B* è lo stimatore corretto ($\tilde{\phi}_{Boot}$)

Shift standardizzato e limiti di controllo

```

shift<-shift*sqrt(1/(1-(mean(ph))^2))
shift
shiftB<-shift*sqrt(1/(1-(mean(phB))^2))
shiftB

```

```

# Valori UCL-LCL
L<-(mean(sigma))*h*sqrt(lambda/(2-lambda))
L
L_B<-(mean(sigmaB))*h*sqrt(lambda/(2-lambda))
L_B

```

Ciclo per il conteggio degli ARL, FCNS e FALL

```

for(j in 1:nsim)
{
  for(k in 1:ncampione)
  {
    if(w[j,k]>L)
    {
      stat[j,k]<-1
    }
  }
  arl[j]<-min(which(stat[j,c((IC+1):ncampione)]==1))
  if(arl[j]==Inf)
  {
    arl[j]=(ncampione-IC)+1
  }
  fcns[j]<-length(which(stat[j,c((IC+1):ncampione)]==0))
  fall[j]<-length(which(stat[j,c(1:IC)]==1))

  for(z in 1:ncampione)
  {
    if(wB[j,z]>L_B)
    {

```

```

        statB[j,z]<-1
    }
}
arlB[j]<-min(which(statB[j,c((IC+1):ncampione)]==1))
if(arlB[j]==Inf)
{
    arlB[j]=(ncampione-IC)+1
}
fcnsB[j]<-length(which(statB[j,c((IC+1):ncampione)]==0))
fallB[j]<-length(which(statB[j,c(1:IC)]==1))
}

```

Analisi conclusive di conteggio e di supporto grafico

```

mean(x)
mean(mu)
mean(muB)

mean(sigma)
mean(sigmaB)

mean(ph)
mean(phB)

# Controllo sui residui
plot(res,type="l",ylim=c(-9,9))
lines(resB,type="l",col=2)
abline(v=IC,col=3)

```

```

mean(res)

mean(resB)

mean(res[1:IC])
mean(resB[1:IC])

mean(res[IC:(ncampione-1)])
mean(resB[IC:(ncampione-1)])

# Primo valore che esce dai limiti
mean(arl)
mean(arlB)
boxplot(arl,arlB,names=c("rl_ML","rl_BOOT"))

# Fuori controllo non segnalati
mean(fcns)
mean(fcnsB)
boxplot(fcns,fcnsB,names=c("fcns","fcnsB"))

# Falsi allarmi
mean(fall)
mean(fallB)
boxplot(fall,fallB,names=c("fall","fallB"))

# Boxplot
boxplot(as.data.frame(w))
abline(h=L,col=2)
abline(h=-L,col=2)
abline(v=IC,col=3)
abline(v=n,col=4)

# "Spaghetti" plot

```

```

plot(ts(w[1,]),ylim=c(min(w),max(w)))
abline(h=L,col=2)
abline(h=-L,col=2)
abline(v=IC,col=1,lty=3)
abline(v=n,col=4)
for(z in 2:nsim)
{
  lines(ts(w[z,]))
}

# Boxplot 2
boxplot(as.data.frame(wB))
abline(h=L,col=2)
abline(h=-L,col=2)
abline(v=IC,col=3)
abline(v=n,col=4)

# "Spaghetti" plot 2
plot(ts(wB[1,]),ylim=c(min(wB),max(wB)))
abline(h=L,col=2)
abline(h=-L,col=2)
abline(v=IC,col=3)
abline(v=n,col=4)
for(z in 2:nsim)
{
  lines(ts(wB[z,]))
}

mean(ph)

```

```
hist(ph)
var2_ph<-sum((ph-mean(ph))^2)/(nsim-1)
var2_ph

mean(phB)
hist(phB)
var2_phB<-sum((phB-mean(phB))^2)/(nsim-1)
var2_phB

plot(ph,type="l",ylim=c(-1,1))
lines(phB,type="l",col=2)
abline(h=min(ph))
abline(h=max(ph))
abline(h=min(phB),col=2)
abline(h=max(phB),col=2)
```

Bibliografia

- Andrews D.W.K. & Chen H.Y. (1994) , Approximate median unbiased estimation of auto regressive models. *Journal of Business & Economic Statistics*, 12, 187-204;
- Andrews D.W.K. (1993) , Exactly median-unbiased estimation of first order autoregressive / unit root models. *Econometrica*, 61, 139-165;
- Apley D. (2002) , Time series control charts in the presence of model uncertainty. *Journal of Manufacturing Science and Engineering*, 124, 891-898;
- Apley D. and Lee H. (2003) , Design of exponentially weighted moving average control charts for autocorrelated processes with model uncertainty. *Technometrics*, 45, 187-198;
- Crowder S.V. (1987) , A simple method for studying run-length distributions of exponentially weighted moving average charts, *Technometrics*, 29, 401-407;
- Crowder S.V. (1989) , Design of exponentially weighted moving average schemes, *Journal of Quality Technology*, 21, 155-162;
- Kilian L. (1998) , Small sample confidence intervals for impulse response functions. *The Review of Economics and Statistics*, 80, 218-230;
- Kim J.H. (2001) , Bootstrap-After-Bootstrap Prediction Intervals for Autoregressive Models;

- Kim J.H. (2003) , Forecasting Autoregressive Time Series with Bias-Corrected Parameter Estimators, *International Journal of Forecasting*, 19, 493-502;
- Kim J.H. (2004) , Bootstrap Prediction Intervals for Autoregression using Asymptotically Mean-Unbiased Parameter Estimators, *International Journal of Forecasting*, 20, 85-97;
- Knoth S. (2005) , Accurate ARL computation for EWMA-S2 control charts, *Statistics and Computing*, 15, 341-352;
- Lucas J.M. and M.S. Saccucci (1990) , Exponentially weighted moving average control schemes: Properties and enhancements, *Technometrics* ,32, 1-12;
- Masarotto G. – G. Capizzi (2007), Bootstrap-based design of residual control charts, Working Paper Series, Department of Statistical Sciences, University of Padua;
- Masarotto G. – G. Capizzi (2007), Practical Design of Generalized Likelihood Ratio Control Charts for Autocorrelated Data, Working Paper Series, Department of Statistical Sciences, University of Padua;
- Masarotto G. (1990) , Bootstrap Prediction Intervals for Autoregression. *International Journal of Forecasting*, 6, 229–329;
- Masarotto G. , (dispense del corso) *Analisi delle serie temporali*;
- Montgomery Douglas C. (2006), *Controllo statistico di qualità* , McGraw-Hill;
- Piccolo D. (1990), *Introduzione all'analisi delle serie storiche* , N.I.S.;
- Roy A. & Fuller W.A. (2001) , Estimation for autoregressive time series with a root near one. *Journal of Business & Economic Statistics*, 19, 482-493;
- Shaman P. & Stine R.A. (1988) , The bias of autoregressive coefficient estimators. *Journal of the American Statistical Association*, 83, 842-848;
- Steiner S. H. (1999) , EWMA control charts with time-varying control limits and fast initial response, *Journal of Quality Technology*, 31, 75-86;

- Stine R.A. & Shaman P. (1989) , A fixed point characterization for bias of autoregressive estimators. *The Annals of Statistics*, 17, 1275-1284;

