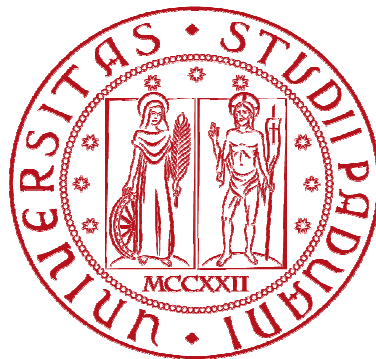


UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI SCIENZE STATISTICHE

**CORSO DI LAUREA IN
STATISTICA, ECONOMIA E FINANZA**



TESI DI LAUREA

**METODI BOOTSTRAP
PER SERIE STORICHE**

RELATORE: Chiar.mo Prof. Luisa Bisaglia

LAUREANDA: Claudia Di Caterina

MATRICOLA N° 600136

ANNO ACCADEMICO 2010-2011

Indice

Introduzione	5
Capitolo 1 – METODI BOOTSTRAP PER SERIE STORICHE	7
1.1 Introduzione	7
1.2 Metodi bootstrap	9
1.2.1 <i>Model-based resampling</i>	9
1.2.2 <i>Block bootstrap</i>	13
1.2.3 <i>Sieve bootstrap</i>	18
Capitolo 2 – BREAK STRUTTURALI	23
2.1 Introduzione	23
2.2 Test per la verifica dell’ipotesi di <i>break</i> strutturale	25
2.2.1 Test per la verifica di un <i>break</i> strutturale ad una data nota	26
2.2.2 Test per la verifica di un <i>break</i> strutturale ad una data ignota	28
Capitolo 3 – APPLICAZIONE DEL BOOTSTRAP	31
3.1 Introduzione	31
3.2 Esperimenti di simulazione Monte Carlo	32
3.3 Applicazione del bootstrap	35
3.4 Conclusioni	41
Appendice	45
A.1 Codice R per gli esperimenti di simulazione Monte Carlo	45
A.2 Codice R per l’applicazione del bootstrap	47
Bibliografia	55

Introduzione

Dopo la pubblicazione nel 1979 dell'articolo di Efron, che propose per primo la metodologia bootstrap basata su tecniche di simulazione e ricampionamento, la decade relativa agli anni Ottanta fu un periodo caratterizzato da intensa ricerca sui metodi bootstrap per dati indipendenti (specialmente i.i.d.). Al contrario, dagli anni Novanta in poi la maggior parte degli studi sono stati rivolti allo sviluppo di nuovi metodi di ricampionamento per dati dipendenti, soprattutto serie storiche. L'applicazione dei metodi bootstrap a serie storiche ci permette di fare inferenza su un parametro ignoto del processo stocastico generatore dei dati, valutando delle proprietà come la distorsione e la varianza di un suo stimatore, e ci consente anche di ricavare una buona approssimazione delle funzioni di distribuzione campionarie di importanti statistiche test, le cui proprietà asintotiche in campioni finiti non sono del tutto soddisfacenti. Naturalmente, nel caso di dati dipendenti, la situazione è più complicata rispetto al caso di dati i.i.d., perché la procedura bootstrap deve essere eseguita in modo tale da catturare correttamente la struttura di dipendenza temporale del processo generatore dei dati.

Nell'ambito delle serie storiche sono stati sviluppati negli ultimi anni molti metodi bootstrap; tra i più utilizzati troviamo il *model-based resampling* (Freedman, 1981), il *block bootstrap* (Künsch, 1989), il *sieve bootstrap* (Swanepoel e van Wyk, 1986), il *phase scrambling* (Theiler *et al.*, 1992), l'*autoregressive-aided periodogram bootstrap* (Kreiss e Paparoditis, 2003), il *local bootstrap* (Paparoditis e Politis, 1999), il *kernel bootstrap* (Dahlhaus e Janas, 1996; Franke ed Härdle, 1992), e l'*ACF bootstrap* basato su un risultato di Ramsey (1974). Nonostante questa grande varietà di metodi oggi disponibili, per alcuni modelli di serie storiche l'applicazione della tecnica bootstrap può comunque non portare a validi risultati. Inferenze ragionevolmente accurate col bootstrap sono particolarmente difficili da ottenere quando si è in presenza di campioni di dimensioni limitate. Conseguentemente, il bootstrap è tuttora un argomento di ricerca attiva, e la classe di modelli per cui esso può efficacemente essere utilizzato è in continuo aumento.

Prendendo atto dell'esistenza di molteplici procedure utilizzate per applicare il bootstrap a dati in forma di serie storiche, in questa tesi concentriamo l'attenzione sui tre metodi più popolari - ovvero il *model-based resampling* o *parametric bootstrap*, il *block bootstrap* ed il *sieve bootstrap* - fornendo un esempio della loro applicazione pratica.

La tesi è suddivisa in tre capitoli: i primi due costituiscono la parte teorica, mentre l'ultimo riporta i risultati di simulazioni e repliche di serie storiche effettuate mediante l'algoritmo bootstrap. Il primo capitolo è dedicato ad una rassegna dei tre metodi bootstrap considerati: viene illustrata la procedura da adottare per ogni metodo, e ne vengono messi in luce vantaggi e svantaggi per differenti tipi di modelli di serie storiche. Il secondo capitolo introduce il concetto di rotture strutturali ed espone dettagliatamente la strategia basata sui test F per individuarne la presenza all'interno di funzioni di regressione temporali, ovvero il test di Quandt (1960) per verificare la stabilità dei coefficienti di regressione ad una data ignota. Questa statistica viene poi ripresa in considerazione nel terzo capitolo, in modo tale da fornire un semplice esempio di applicazione della tecnica bootstrap a serie storiche simulate. Per mezzo di esperimenti iniziali di simulazione Monte Carlo viene messo in evidenza come l'ampiezza (detta anche dimensione o *size*, in inglese) effettiva del test di Quandt risulti differente da quella nominale del 5%, quando calcolato su serie storiche simulate AR(1) e AR(2) e confrontato con gli usuali valori critici ricavati dalla distribuzione asintotica del test. Questo avviene in particolar modo per alcuni valori dei parametri autoregressivi. Si cerca quindi di verificare se, impiegando i tre metodi bootstrap presi in esame al fine di ottenere una valida approssimazione della distribuzione campionaria della statistica test, la *size* effettiva stimata si avvicini o meno a quella nominale. Nonostante il numero di simulazioni con cui viene effettuata questa analisi sia considerevolmente inferiore a quello adeguato, i risultati conseguiti danno evidenza di una significativa riduzione della distorsione nella stima della dimensione effettiva del test di Quandt, confermando quindi gli attesi miglioramenti apportati dal bootstrap nella procedura di stima.

Capitolo 1

METODI BOOTSTRAP PER SERIE STORICHE

1.1 Introduzione

L'idea base del bootstrap fu introdotta nel 1979 da Bradley Efron, che propose una metodologia basata su calcoli informatici finalizzati alla stima dello scarto quadratico medio di uno stimatore $\hat{\theta}$. Il bootstrap è quindi una tecnica statistica *computer intensive* che, infatti, deve la sua popolarità allo sviluppo di potenti calcolatori elettronici dal semplice utilizzo.

Originariamente i metodi bootstrap venivano soprattutto applicati per ottenere misure dell'accuratezza di una statistica d'interesse, ma al giorno d'oggi sono utilizzati per vari scopi, tra cui la costruzione di intervalli di confidenza, la verifica di ipotesi e la stima della distribuzione di test statistici.

La logica su cui poggia il bootstrap per dati indipendenti ed identicamente distribuiti è molto semplice e può essere illustrata come segue: sia x_1, \dots, x_n un campione iniziale estratto da una variabile casuale X , e supponiamo di essere interessati ad un parametro θ tale che $\hat{\theta} = T(x_1, \dots, x_n)$, dove $T(\cdot)$ è una statistica; dal campione iniziale si estraggono con reinserimento altri B campioni $x_{1,b}^*, \dots, x_{n,b}^*$, con $b = 1, \dots, B$. Quindi su ciascuno di questi campioni bootstrap si stima, con lo stesso stimatore $\hat{\theta}_b^*$, il parametro di interesse. L'insieme di $\hat{\theta}_b^*$ è chiamato distribuzione bootstrap di $\hat{\theta}$, e può essere usato per gli scopi già menzionati. In Figura 1.1 vi è la rappresentazione grafica della procedura bootstrap ideata da Efron (1979) per campioni casuali semplici.

Sfortunatamente le cose non sono così semplici con dati non i.i.d, come, per esempio, le serie storiche. Questa tecnica, infatti, distruggerebbe completamente la struttura di dipendenza temporale di qualunque sequenza di dati autocorrelati. Pertanto, altri metodi di ricampionamento devono essere utilizzati, in modo da riprodurre la struttura di dipendenza temporale ed ottenere delle

nuove serie le cui caratteristiche di dipendenza tra le osservazioni siano simili a quelle della serie storica osservata. Li e Maddala (1996) hanno discusso le difficoltà incontrate nell'applicare l'algoritmo bootstrap a dati in forma di serie storiche. Queste, infatti, violano l'originaria assunzione di dati indipendenti ed identicamente distribuiti del bootstrap ideato da Efron, quindi le osservazioni non sono intercambiabili. Più recentemente, Bühlmann (2002) ha passato in rassegna e messo a confronto alcuni metodi bootstrap per serie storiche, soffermandosi sugli aspetti teorici delle procedure e sulle loro *performance* in campioni finiti di dati. Politis (2003) ha anche mostrato che le tecniche bootstrap permettono di stimare consistentemente la varianza della funzione di autocorrelazione campionaria $\hat{\rho}$, risultato importante perché le formule di Bartlett sono disponibili solo per processi lineari. Inoltre, sotto l'assunzione di stazionarietà del processo, Politis ha messo in evidenza che il bootstrap permette stime più accurate della distribuzione della media campionaria. Ulteriori recenti studi sui metodi bootstrap per dati dipendenti includono, tra gli altri, Horowitz (2003) ed Härdle, Horowitz e Kreiss (2003).

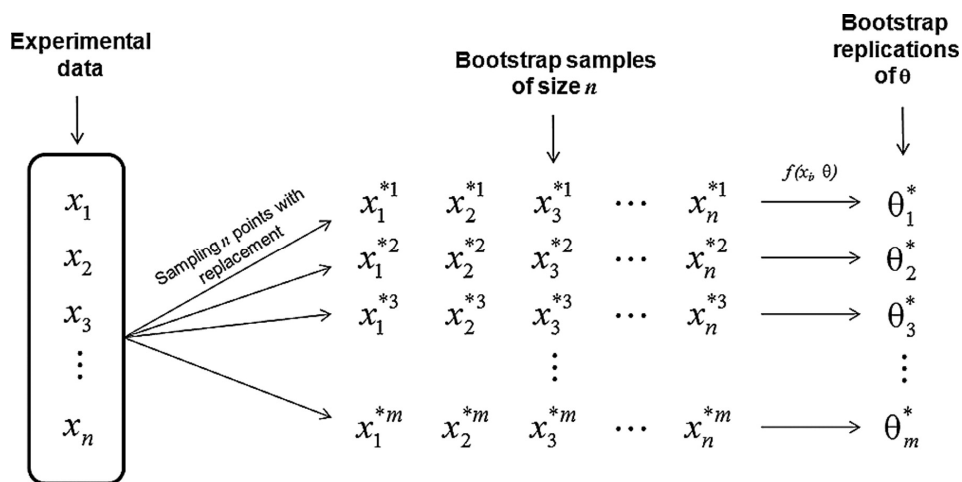


Figura 1.1: Rappresentazione grafica della procedura bootstrap per dati i.i.d.

(fonte: Hernandez e Mulla, 2008).

Nonostante la vasta letteratura riguardante le tecniche bootstrap per serie storiche, il problema del ricampionamento di sequenze di dati autocorrelati è ancora aperto, poiché tali tecniche non sono sempre soddisfacenti, in parti-

colar modo nell'ambito di serie storiche che presentano una dipendenza temporale di lungo periodo.

Nelle prossime sezioni di questo capitolo ci soffermeremo sui tre principali metodi bootstrap applicati alle serie storiche: il *model-based resampling*, il *block bootstrap* ed il *sieve bootstrap*.

1.2 Metodi bootstrap

Supponiamo di disporre di una serie storica Y_t per $t = 1, \dots, T$, generata da un processo stocastico ignoto, e di essere interessati, per effettuare inferenza su un determinato parametro incognito del processo, alla distribuzione empirica di un certo stimatore $\hat{\varphi}$ funzione delle nostre osservazioni, ovvero tale che $\hat{\varphi} = h_T(Y_1, \dots, Y_T)$.

1.2.1 *Model-based resampling*

Questo approccio per regressioni e autoregressioni lineari fu introdotto e studiato, quasi immediatamente dopo il lavoro di Efron del 1979, da Freedman (1981, 1984) e da Efron e Tibshirani (1986, 1993); esso fa parte della classe dei metodi di ricampionamento basati sui residui (Berkowitz e Kilian, 1996) e viene anche chiamato *parametric bootstrap*, o bootstrap parametrico. Il metodo è applicabile se si assume che il vero meccanismo generatore dei dati è un processo ARMA stazionario ed invertibile, di ordine finito, e con innovazioni i.i.d. (Efron e Tibshirani, 1986). In questo caso, infatti, è possibile stimare i residui del processo. In realtà, Lahiri nel 1992 ha sviluppato un bootstrap parametrico per processi autoregressivi instabili ed esplosivi, ma la procedura da effettuare in tal caso si complica notevolmente, quindi in questa sede ci occuperemo solamente del caso stazionario. Sotto l'assunzione di corretta specificazione del modello, i residui sono indipendenti ed identicamente distribuiti, e ciò rende possibile applicare a questi il classico bootstrap i.i.d.. Le pseudo-serie boot-

strap sono quindi costruite utilizzando i residui ricampionati e i parametri stimati. Le fasi principali del *parametric bootstrap* sono le seguenti:

1. Adattare un modello appropriato ai dati, per esempio un $AR(p)$, scegliendo il miglior modello in base ad un certo criterio (AIC, ecc.).
2. Stimare i residui in base al modello scelto.
3. Applicare ai residui il bootstrap i.i.d..
4. Usare i parametri stimati e i residui bootstrap ricampionati per costruire le serie bootstrap che possiedono proprietà simili alla serie originale.

La letteratura è ricca di lavori relativi al *model-based bootstrap*. Tra i più importanti ricordiamo Swanepoel e van Wyk (1986) che hanno utilizzato questo metodo per costruire bande di confidenza per la densità spettrale, e Bose (1988) che ha mostrato che la distribuzione bootstrap delle stime OLS dei parametri autoregressivi risulta più accurata rispetto all'approssimazione normale.

Soffermiamoci sulla procedura del *parametric bootstrap* nel caso di un modello $AR(p)$. Il primo passo consiste nell'identificazione, a partire dalla serie storica di cui disponiamo, del vero processo generatore dei dati e, quindi, nella specificazione dell'ordine del modello AR (parametro p) tale che:

$$(Y_t - \mu_Y) - \sum_{i=1}^p \phi_i (Y_{t-i} - \mu_Y) = \varepsilon_t \quad \text{ovvero} \quad \phi(B)(Y_t - \mu_Y) = \varepsilon_t,$$

$$\text{per } t = 1, \dots, T,$$

$$\text{con } \mu_Y = E(Y_t), \quad \varepsilon_t \sim i.i.d., \quad E(\varepsilon_t) = 0, \quad E(\varepsilon_t^2) < \infty,$$

$$\text{e } \phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$$

polinomio invertibile nell'operatore ritardo B , tale che $Var(Y_t) < \infty$.

Successivamente si passa alla fase di stima dei parametri, che può essere effettuata con le equazioni di Yule-Walker, col metodo dei minimi quadrati linea-

ri, oppure col metodo della massima verosimiglianza. La stima della media del processo μ_Y è invece ottenuta utilizzando la media campionaria:

$$\widehat{\mu}_Y = \frac{1}{T} \sum_{t=1}^T Y_t.$$

Naturalmente in questo contesto dobbiamo cercare di stimare i parametri ignoti nel modo più preciso possibile, dal momento che queste stime saranno poi utilizzate per definire il processo generatore dei dati bootstrap.

Una volta ottenute le stime puntuali del vettore dei parametri, $(\widehat{\mu}_Y, \widehat{\phi}_1, \dots, \widehat{\phi}_p)$, possiamo generare innovazioni bootstrap \mathcal{E}_t^* mediante il ricampionamento con reinserimento dei residui stimati centrati $\widetilde{\mathcal{E}}_t$, che si ottengono nel seguente modo:

$$\widehat{\mathcal{E}}_t = \left(Y_t - \widehat{\mu}_Y \right) - \sum_{i=1}^p \widehat{\phi}_i \left(Y_{t-i} - \widehat{\mu}_Y \right) \text{ oppure } \widehat{\mathcal{E}}_t = \widehat{\phi}(B) \left(Y_t - \widehat{\mu}_Y \right),$$

per $t = p+1, \dots, T$.

Da questi ricaviamo i residui centrati $\widetilde{\mathcal{E}}_t = \widehat{\mathcal{E}}_t - \frac{1}{T-p} \sum_{t=p+1}^T \widehat{\mathcal{E}}_t$, come suggerito da Lahiri (1992).

Ricampionare significa estrarre gli \mathcal{E}_t^* dalla funzione di distribuzione empirica di $\widetilde{\mathcal{E}}_t$, che assegna probabilità $\frac{1}{T-p}$ ad ognuno degli $\widetilde{\mathcal{E}}_t$.

Se, invece, è plausibile assumere che i residui siano Normalmente distribuiti, si può applicare un metodo completamente parametrico, generando le innovazioni bootstrap da una distribuzione Normale con media zero e varianza stimata da $\widehat{\sigma}^2 = \widehat{\text{var}}(\widetilde{\mathcal{E}}_t)$.

A questo punto, come suggeriscono Berkowitz e Kilian (1996), scegliamo un numero intero positivo elevato τ e poniamo $Y_t^* = \widehat{\mu}_Y$ per $t < -\tau$. I valori successivi di Y_t^* sono invece ottenuti in questo modo:

$$Y_t^* - \widehat{\mu}_Y = \sum_{i=1}^p \widehat{\phi}_i (Y_{t-i}^* - \widehat{\mu}_Y) + \varepsilon_t^* \text{ oppure } Y_t^* - \widehat{\mu}_Y = \widehat{\phi}(B)^{-1} \varepsilon_t^*$$

per $t = -\tau, \dots, T$.

Abbiamo così costruito una pseudo-serie più lunga di quella osservata, in modo da poter poi eliminare il segmento iniziale dei dati simulati e conservare solo gli ultimi T valori di Y_t^* (effettuando il cosiddetto *burn-in* per processi stazionari), al fine di ottenere la nostra serie bootstrap $\mathbf{Y}_{t,b}^* = (Y_{1,b}^*, \dots, Y_{T,b}^*)$. Calcoliamo quindi la statistica di interesse $\widehat{\phi}$ su questa serie:

$$\widehat{\phi}_b^* = h_T(Y_{1,b}^*, \dots, Y_{T,b}^*).$$

A questo punto, ripetiamo la procedura illustrata (dalla fase di ricampionamento dei residui in poi) B volte, con B numero intero positivo molto grande, calcolando per ogni nuovo campione bootstrap $\widehat{\phi}_b^*$, in modo tale da poter ricavare la distribuzione empirica della statistica d'interesse.

La procedura per applicare il bootstrap ad un processo autoregressivo a media mobile ARMA (p, q) è molto simile a quella appena esposta (Chatterjee, 1986, e Kreiss e Franke, 1989). Il modello stimato può essere scritto nella forma:

$$(Y_t - \mu_Y) - \sum_{i=1}^p \phi_i (Y_{t-i} - \mu_Y) = \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j} \text{ ovvero } \phi(B)(Y_t - \mu_Y) = \theta(B)\varepsilon_t$$

per $t = 1, \dots, T$,

$$\text{con } \mu_y = E(Y_t), \varepsilon_t \sim i.i.d., E(\varepsilon_t) = 0, E(\varepsilon_t^2) < \infty,$$

$$\text{e } \phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p) \text{ e } \theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$$

polinomi invertibili nell'operatore ritardo B , tali che $Var(Y_t) < \infty$. Ciò che cambia per un modello ARMA(p, q) rispetto alla procedura per un autoregressivo puro è la formula per stimare i residui, che diventa:

$$\hat{\varepsilon}_t = \sum_{j=1}^t \tilde{\theta}_{j-1} \left(- \sum_{k=0}^p \hat{\phi}_k X_{t+1-j-k} \right),$$

dove $\sum_{j=0}^{\infty} \tilde{\theta}_j B^j = (\hat{\theta}(B))^{-1}$, $\hat{\phi}_k$ per $k = 1, \dots, p$, e $\hat{\theta}_j$ per $j = 1, \dots, q$, sono stimatori consistenti dei parametri autoregressivi e a media mobile, e $\hat{\phi}_0 = -1$.

Questo metodo parametrico è appropriato per una ristretta classe di processi, ovvero quelli lineari, e dà buoni risultati solamente se il modello è correttamente specificato, altrimenti le serie bootstrap generate non hanno le stesse proprietà del processo osservato (Davison ed Hinkley, 1997).

1.2.2 *Block bootstrap*

Il bootstrap a blocchi introdotto da Künsch nel 1989 è senz'altro il metodo più conosciuto ed utilizzato per implementare il bootstrap con dati di serie storiche (Härdle, Horowitz e Kreiss, 2001). L'idea centrale di questo metodo è la stessa del bootstrap per dati i.i.d.: entrambe le procedure si basano, infatti, sull'estrazione casuale con reinserimento delle osservazioni della serie oggetto di analisi. Nel *block bootstrap*, tuttavia, vengono estratti blocchi di l osservazioni consecutive, non singole osservazioni; questo permette di preservare la struttura di dipendenza esistente tra osservazioni vicine nell'originale serie storica, almeno entro ogni blocco.

È stato mostrato che tale approccio funziona per gran parte dei processi stazionari (Künsch, 1989) e, pur non raggiungendo il livello di precisione ottenuto dal bootstrap per dati i.i.d., consegue risultati ragionevolmente soddisfacenti quando applicato a serie storiche che presentano una dipendenza temporale di breve periodo (Liu e Singh, 1992). Nessuna assunzione viene fatta sulla struttura del processo generatore dei dati, e questo fa del *block bootstrap* un metodo non parametrico che, a differenza del *model-based resampling*, non può essere affetto da problemi causati da una non corretta specificazione del modello, e presenta, inoltre, il vantaggio di non richiedere stime di parametri ignoti che definiscono la struttura del processo.

Esistono diverse versioni del *block bootstrap* che si differenziano in base alla scelta di blocchi sovrapposti (Hall 1985, Künsch 1989, Politis e Romano 1993) o non sovrapposti (Hall 1985, Carlstein 1986), e in base alla lunghezza l dei blocchi, che può essere variabile (Politis e Romano, 1993) o fissa (Bühlmann, 2002). Tuttavia, è importante evidenziare il fatto che tutti questi metodi hanno in comune il ricampionamento di blocchi di osservazioni consecutive. Nel caso di blocchi non sovrapposti, sviluppato da Carlstein nel 1986, la serie storica osservata viene suddivisa in c blocchi B_1, \dots, B_{T-l} così costruiti:

$$B_k = \left(Y_{(k-1)l+1}, \dots, Y_{kl} \right), \text{ con } k = 1, \dots, c,$$

dove $lc \leq T$.

Tuttavia, studi teorici e varie simulazioni suggeriscono che l'approccio migliore sia quello di utilizzare blocchi sovrapposti di lunghezza fissa (Hall *et al.* 1995, e Lahiri 1999). Questa variante è detta *moving-block bootstrap* (o meno frequentemente, *blockwise bootstrap*) e la procedura è illustrata di seguito.

Costruiamo un insieme di $T-l+1$ blocchi sovrapposti di osservazioni consecutive in questo modo:

$$B_1 = (Y_1, \dots, Y_l), B_2 = (Y_2, \dots, Y_{l+1}), \dots, B_k = (Y_k, \dots, Y_{l+k-1}), \dots,$$

$$B_{T-l+1} = (Y_{T-l+1}, \dots, Y_T), \quad l \in \mathbb{N}.$$

Quindi, otteniamo la nostra serie storica bootstrap $Y_{t,b}^*$ di lunghezza T , ricampionando in modo casuale e con reinserimento $\frac{T}{l}$ blocchi tra quelli creati, ponendoli uno di seguito all'altro nell'ordine di estrazione. Se $\frac{T}{l}$ non è un numero intero, uno o più blocchi dovranno essere troncati per formare un campione di numerosità T (MacKinnon, 2006).

Ripetiamo questa procedura B volte (con B numero intero positivo molto grande), calcolando per ogni nuovo campione bootstrap la nostra statistica d'interesse $\hat{\varphi}_b^*$.

La differenza tra blocchi sovrapposti e blocchi non sovrapposti, utilizzati in diverse versioni del *block bootstrap* per serie storiche, è illustrata di seguito in Figura 1.2:

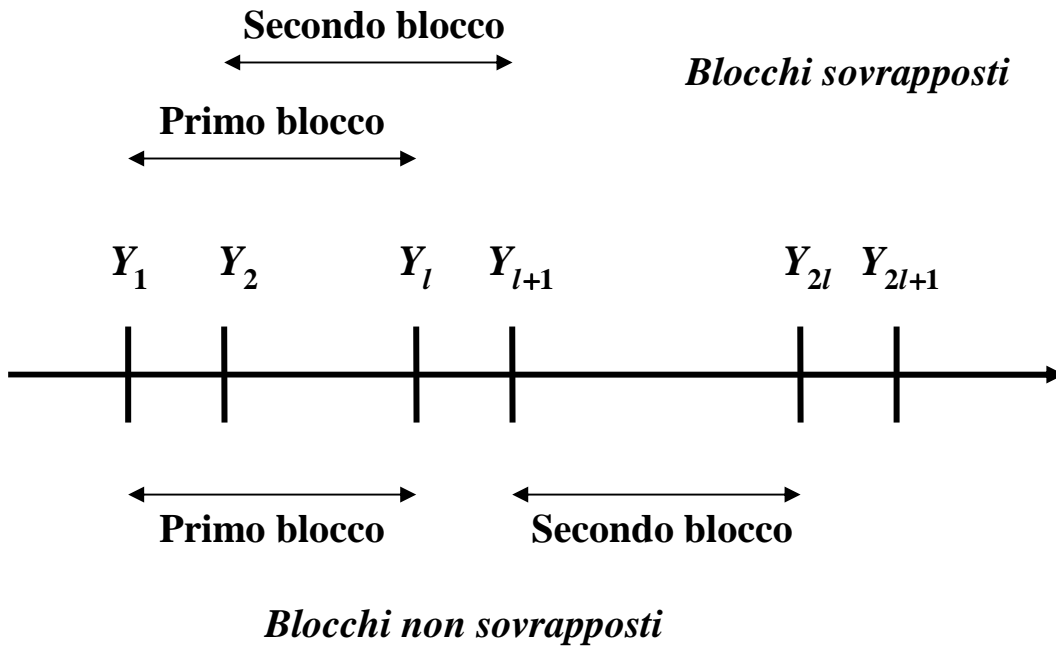


Figura 1.2: blocchi sovrapposti e blocchi non sovrapposti usati nel block bootstrap.

È chiaro che la scelta della lunghezza dei blocchi l risulta di estrema importanza, sia nel caso di blocchi sovrapposti sia in quello di blocchi non sovrapposti: se l è fissata troppo piccola, i campioni bootstrap non riescono a riprodurre adeguatamente le relazioni di dipendenza temporale dei dati originali, perché queste sono interrotte ogni volta che un blocco finisce ed un altro inizia, pertanto l'accuratezza complessiva del *block bootstrap* diminuisce. D'altra parte, se l è troppo grande, i campioni bootstrap tendono ad essere eccessivamente influenzati dalle caratteristiche casuali del vero campione, ed essendoci pochi blocchi, questi pseudo-campioni sono anche troppo simili tra loro (Berkowitz e Kilian, 1996 e Davidson e MacKinnon, 2005). Inoltre, la scelta della dimensione dei blocchi presenta un *trade-off* tra varianza e distorsione: blocchi più lunghi aumentano la varianza dello stimatore, ma riducono la sua distorsione, essendoci meno blocchi da ricampionare (Lahiri, 2003).

La lunghezza ottimale dei blocchi dipende almeno da tre fattori: il processo generatore dei dati, la statistica d'interesse che deve essere calcolata in ogni campione bootstrap e lo scopo per cui il bootstrap viene applicato, ovvero per fare inferenza sulla distorsione o sulla varianza di uno stimatore, oppure per stimare la sua distribuzione empirica (Bühlmann, 2002). Quando, per esempio, l'obiettivo è stimare un momento o la funzione di distribuzione di uno stimatore, la lunghezza asintotica ottimale l è quella che minimizza l'errore quadratico medio asintotico (AMSE) dello stimatore bootstrap (Carlstein 1986, e Ouyse 2008).

Da un punto di vista asintotico, l dovrebbe crescere, ma non troppo velocemente, al crescere della lunghezza T della serie storica osservata, in modo da ottenere stimatori bootstrap di momenti e funzioni di distribuzione che siano consistenti (Bühlmann, 2002). Alcuni studi evidenziano, infatti, che la lunghezza ottimale dei blocchi è spesso proporzionale a $T^{1/3}$, ma con un fattore di proporzionalità che è praticamente sconosciuto, cioè $l_{opt} = \text{cost} \cdot T^{1/3}$ (Härdle, Horowitz e Kreiss 2003, e Bühlmann 2002, sulla base dei risultati di Carlstein 1986, Künsch 1989 ed Hall *et al.* 1995). In generale, solo quando la numerosità cam-

pionaria è molto grande è possibile scegliere l in modo che i blocchi non siano né troppo lunghi né troppo corti, e quindi il processo generatore dei dati bootstrap fornisca una valida approssimazione del vero processo generatore dei dati (Davidson e MacKinnon, 2005).

Ovviando a questo problema, Politis e Romano nel 1994 hanno proposto una versione del bootstrap a blocchi in cui la lunghezza di ogni blocco è casuale e generata da una distribuzione geometrica di parametro p , anziché fissata uguale per tutti i blocchi. Questo approccio è detto *stationary bootstrap*, proprio perché le serie storiche bootstrap risultanti sono sicuramente stazionarie (vedi proposizione 1, Politis e Romano 1994), mentre ciò non accade nel *moving-block bootstrap*. Lo *stationary bootstrap*, infatti, ottiene la stazionarietà delle pseudo-serie ricampionate, perché considera le singole osservazioni della serie originale disposte circolarmente, ovvero Y_1 segue Y_T , in modo tale da assegnare ad ogni osservazione la stessa probabilità di essere estratta durante il ricampionamento dei blocchi (per una spiegazione più dettagliata, vedi Politis e Romano 1994). Questo metodo ovvia anche ad un ulteriore inconveniente del *moving-block bootstrap*, ossia la distorsione delle stime bootstrap della media e della varianza del processo. Inoltre, risultati di simulazione dimostrano che la scelta del parametro p è meno problematica di quella della lunghezza l dei blocchi, nel senso che la scelta di un valore sbagliato provoca un impatto meno dannoso sul risultato finale.

Sono state sviluppate anche altre variazioni rispetto alla procedura classica del *moving-block bootstrap*, per tentare di migliorarne la precisione riducendo l'influenza dei "punti di discontinuità" che si vengono a creare tra un blocco e l'altro nelle pseudo-serie bootstrap (tra le più popolari citiamo il *matched-block bootstrap* introdotto da Carlstein *et al.* nel 1998, ed il *tapered block bootstrap* ideato da Papanoditis e Politis nel 2001). Ma, come già accennato all'inizio di questa sottosezione, ad oggi il *moving-block bootstrap* sembra essere il metodo migliore tra le varie versioni del bootstrap a blocchi.

Un indubbio vantaggio dei metodi basati sul ricampionamento di blocchi di osservazioni consecutive consiste nella loro semplicità concettuale. Tuttavia, il loro più grande problema è che spesso non funzionano molto bene. Da un

punto di vista teorico, il fatto è che, anche quando il *block bootstrap* offre accuratezza di ordine superiore rispetto ai metodi asintotici, lo fa solo in misura modesta, quindi le sue proprietà non risultano del tutto soddisfacenti (Davidson e MacKinnon, 2005). Inoltre, in generale, non ha rendimenti molto buoni quando applicato a campioni finiti (MacKinnon, 2002 e 2006).

1.2.3 *Sieve bootstrap*

Questo metodo, definito sia semiparametrico (Davidson e MacKinnon, 2005) sia non parametrico (Berkowitz e Kilian, 1996), fu proposto da Swanepoel e van Wyk (1986) come mezzo per ottenere stime consistenti della varianza e della distribuzione di statistiche associate a dati dipendenti, e successivamente rigorosamente investigato, tra gli altri, da Kreiss (1988 e 1992), Paparoditis (1992) e Bühlmann (1997 e 1998). Il *sieve bootstrap* (o *AR-sieve bootstrap*, Bühlmann 2002) offre, sotto certe condizioni, un sostanziale miglioramento rispetto alla *performance* del bootstrap a blocchi, e si basa sul fatto che qualsiasi processo stocastico lineare ed invertibile può essere approssimato da un processo autoregressivo di ordine infinito [AR(∞)]. Quindi, se la serie storica osservata di cui disponiamo può essere considerata una realizzazione finita di un processo stocastico che possiede le suddette proprietà, allora è possibile ricondurla alla forma:

$$Y_t - \mu_Y = \sum_{i=1}^{\infty} \phi_i (Y_{t-i} - \mu_Y) + \varepsilon_t, \quad \text{per } t = 1, \dots, T,$$

dove $\mu_Y = E(Y_t)$, $\varepsilon_t \sim i.i.d.$, $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) < \infty$, e $\sum_{i=1}^{\infty} \phi_i^2 < \infty$.

La procedura illustrata da Bühlmann (2002) inizia adattando alla nostra serie storica un modello AR(p) di questo tipo:

$$Y_t - \mu_Y = \sum_{i=1}^p \phi_i (Y_{t-i} - \mu_Y) + \varepsilon_t, \quad \text{per } t = 1, \dots, T,$$

dove μ_Y ed ε_t soddisfano le stesse assunzioni scritte sopra. A questo punto scegliamo un ordine autoregressivo \hat{p} , avvalendoci del criterio d'informazione automatica di Akaike (AIC) o di un'analisi sequenziale (*sequential testing*, MacKinnon 2006).

Da qui in poi la procedura è molto simile a quella vista per il *model-based resampling* (il ricampionamento è infatti sempre basato sui residui). Stimo μ_Y con la media campionaria $\hat{\mu}_Y = \frac{1}{T} \sum_{t=1}^T Y_t$, e ottengo il vettore $(\hat{\phi}_1, \dots, \hat{\phi}_{\hat{p}})$ utilizzando le equazioni di Yule-Walker (come suggerito da Bühlmann, 1997 e 2002), il metodo della massima verosimiglianza, oppure il metodo dei minimi quadrati lineari (per maggiori dettagli sull'utilizzo di questi metodi, vedi Brockwell e Davis 1998, o Shumway e Stoffer 2000).

Possiamo quindi passare al ricampionamento dei residui centrati in zero $\tilde{\varepsilon}_t$, in modo da generare le nostre innovazioni bootstrap ε_t^* , come visto in precedenza:

$$\hat{\varepsilon}_t = \left(Y_t - \hat{\mu}_Y \right) - \sum_{i=1}^{\hat{p}} \hat{\phi}_i \left(Y_{t-i} - \hat{\mu}_Y \right) \quad \text{e} \quad \tilde{\varepsilon}_t = \hat{\varepsilon}_t - \frac{1}{T - \hat{p}} \sum_{i=\hat{p}+1}^T \hat{\varepsilon}_i,$$

per $t = \hat{p} + 1, \dots, T$.

Si può anche scrivere, come suggerisce Bühlmann (2002):

$$\varepsilon_t^* \sim \widehat{F}_{\tilde{\varepsilon}},$$

$$\text{dove } \widehat{F}_{\tilde{\varepsilon}}(x) = \mathbb{P}(\tilde{\varepsilon}_t \leq x) = (T - \hat{p})^{-1} \sum_{t=\hat{p}+1}^T I_{[\tilde{\varepsilon}_t \leq x]},$$

che equivale ad estrarre casualmente con reinserimento le innovazioni ε_t^* dai residui centrati $\tilde{\varepsilon}_t$, assegnando ad ognuno probabilità di essere estratto pari a

$$\frac{1}{T - \hat{p}}.$$

Dopo aver posto $Y_t^* = \widehat{\mu}_Y$ per $t = -\tau, \dots, -\tau + \widehat{p} - 1$, con τ numero grande, intero e positivo, generiamo i nostri dati bootstrap nel seguente modo:

$$Y_t^* - \widehat{\mu}_Y = \sum_{i=1}^{\widehat{p}} \widehat{\phi}_i (Y_{t-i}^* - \widehat{\mu}_Y) + \varepsilon_t^* \quad \text{per } t = -\tau + \widehat{p}, \dots, 0, 1, \dots, T,$$

conservando alla fine solo gli ultimi T valori Y_1^*, \dots, Y_T^* , come suggerito da Bühlmann (2002).

A questo punto possiamo ripetere la stessa procedura B volte, creando molti altri campioni bootstrap e calcolando per ognuno la nostra statistica di interesse $\widehat{\varphi}^*$.

Il *sieve bootstrap* (che in inglese significa “bootstrap setaccio”) si chiama così proprio perché il modello $AR(\infty)$ viene approssimato da un modello $AR(p)$, il cui ordine $p = p(T)$ dipende dalla numerosità campionaria T , e più precisamente $p(T)$ deve crescere, ma non troppo velocemente, con la lunghezza della serie storica oggetto di analisi, ovvero $p(T) \rightarrow \infty$, con $p(T) = o(T)$, quando $T \rightarrow \infty$ (Bühlmann, 1997). Analogamente al problema di scegliere una lunghezza dei blocchi ottimale nel *block bootstrap*, l’ordine autoregressivo ottimale p , generalmente, dipende dal vero processo stocastico sottostante, dalla statistica d’interesse che si vuole investigare e dal motivo per cui il bootstrap è utilizzato (se per stimare uno dei momenti o la funzione di distribuzione dello stimatore bootstrap). Tuttavia, a differenza della lunghezza dei blocchi, la selezione di un certo ordine autoregressivo può essere facilmente testata attraverso procedure diagnostiche per valutare la bontà del modello adattato ai dati, incluse quelle riguardanti i residui AR. Inoltre, a patto che l’ordine p venga scelto in modo ragionevole, l’esperienza empirica dimostra che la scelta di un particolare ordine autoregressivo piuttosto di un altro non influisce più di tanto sui risultati ottenuti applicando il *sieve bootstrap* (Bühlmann, 2002).

Affinché questo metodo possa essere applicato, il processo generatore dei dati (DGP) deve soddisfare delle assunzioni a priori più stringenti rispetto a quelle necessarie per applicare il *block bootstrap*, e proprio perché il *sieve bootstrap* impone maggiore struttura al DGP, ci si aspetta che consegua risultati migliori del bootstrap a blocchi. In effetti, studi pratici e teorici dimostrano che, entro la classe delle serie storiche lineari ed invertibili, solitamente il *sieve bootstrap* fornisce prestazioni migliori del *moving-block bootstrap*, in particolar modo quando i ricercatori hanno informazioni preliminari riguardo alle proprietà della serie storica presa in esame (Bühlmann, 1997 e 2002). Inoltre, Bühlmann (1997), Kreiss (1999 e 2000) e Paparoditis (1996) hanno mostrato che, sotto appropriate condizioni, il *sieve bootstrap* riesce quasi a raggiungere il livello di precisione del bootstrap i.i.d., riuscendo a stimare consistentemente le distribuzioni empiriche di medie campionarie, autocovarianze e autocorrelazioni campionarie, e coefficienti autoregressivi.

Rispetto al bootstrap parametrico, il *sieve bootstrap* rappresenta un approccio più realistico: nella pratica non è mai possibile sapere con certezza se la struttura stocastica sottostante una data serie storica corrisponda esattamente ad un certo modello ARMA di ordine finito, quindi l'assunzione di corretta specificazione del modello non è mai realmente verificabile. Inoltre, la rappresentazione $AR(\infty)$ su cui si basa il *sieve bootstrap* include anche l'importante classe dei modelli $ARMA(p,q)$ con polinomio a media mobile invertibile (Bühlmann, 2002).

Naturalmente ci sono pure molti processi non rappresentabili come $AR(\infty)$, per esempio i processi autoregressivi non lineari, ed in generale è stato provato che un modello $AR(p)$, con un ragionevole valore di p , non fornisce una buona approssimazione di ogni processo stocastico generatore di serie storiche realmente riscontrate nella pratica (Davidson e MacKinnon, 2005). Sfortunatamente, testare la linearità del vero processo o la possibilità di rappresentarlo con un $AR(\infty)$ è una questione molto delicata. Bickel e Bühlmann (1997) hanno mostrato la particolare difficoltà che si incontra nello stabilire, per un certo insieme di dati, se l'applicazione del *sieve bootstrap* sia consigliabile o meno. Inoltre, soprattutto in econometria, questo metodo bootstrap non è

utilizzato quanto quello a blocchi, perché il fatto di non ammettere errori eteroschedastici comporta una limitazione difficile da superare (Davidson e MacKinnon, 2005).

Molti studi sono stati sviluppati nel corso degli anni per cercare di applicare il *sieve bootstrap* ad una più ampia classe di processi stocastici: Bühlmann (1998) l'ha utilizzato su serie storiche dotate di trend deterministico, più di recente Kapetanios e Psaradakis (2006) e Franco e Reisen (2007) si sono concentrati sui risultati ottenuti quando questo metodo di ricampionamento viene utilizzato nell'ambito dei processi lineari a memoria lunga, e Poskitt (2007) ha investigato le conseguenze della sua applicazione sotto condizioni di regolarità sufficientemente generali da ammettere anche processi non invertibili.

Capitolo 2

***BREAK* STRUTTURALI**

In questo capitolo introduciamo il concetto di *break* strutturali e ci soffermiamo su una delle strategie più comuni per verificare l'ipotesi di presenza di cambiamenti strutturali (o di regime) all'interno di una data serie storica, ovvero il test di Quandt. Questa statistica verrà ripresa in considerazione nel prossimo capitolo, in modo da fornire un esempio pratico dell'applicazione della metodologia bootstrap alle serie storiche.

2.1 Introduzione

Le rotture strutturali (così chiamate in analogia con il termine inglese *breaks*) possono scaturire o da un netto cambiamento nei coefficienti della funzione di regressione temporale ad una data precisa, o da una graduale evoluzione dei coefficienti nel corso di un periodo di tempo più lungo (Stock e Watson, 2005). Queste rotture, che si verificano realmente nelle serie storiche osservate, possono essere causate da vari fattori, come ad esempio cambiamenti nella politica economica, cambiamenti strutturali nell'economia di un Paese, o un'invenzione che trasforma profondamente una specifica industria. Un esempio reale di rottura strutturale è dato dalla serie storica del tasso italiano di irregolarità (rapporto tra il numero di unità di lavoro irregolari e quelle totali), rappresentata in Figura 2.1: come si può notare, la più evidente rottura strutturale si è verificata nel 2002, con effetti fino al 2003, in corrispondenza delle legalizzazioni di più di 600.000 immigrati irregolari (*lavoce.info*).

Una possibile fonte di rotture strutturali nette nei dati macroeconomici è costituita dai cambiamenti radicali nella politica macroeconomica. Per esempio, il collasso del sistema di Bretton Woods di tassi di cambio fissi, avvenuto nel 1972, ha causato una rottura netta nel comportamento del tasso di cambio dollaro/sterlina.

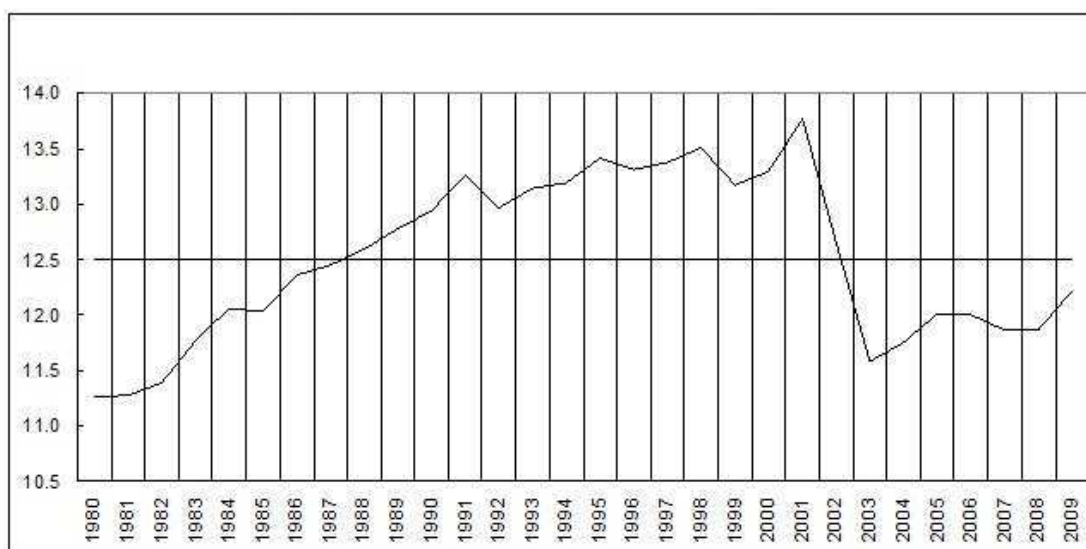


Figura 2.1: serie storica del tasso italiano di irregolarità nel periodo 1980-2009

(fonte: ISTAT).

Le rotture strutturali possono anche verificarsi più lentamente, man mano che la serie storica si evolve nel tempo. Ad esempio, questo tipo di cambiamenti può essere determinato da una lenta evoluzione della politica economica e da mutamenti continui nella struttura dell'economia.

Se queste rotture si verificano durante il periodo campionario, allora il modello di regressione temporale che le ignori può fornire una base fuorviante per l'inferenza e per le previsioni: le stime OLS per l'intero campione, infatti, stimeranno una relazione valida "in media", nel senso che la stima combinerà i due differenti periodi (vedi Figura 2.2). A seconda della posizione e dell'ampiezza della rottura, la funzione di regressione temporale "media" può essere molto diversa dalla vera funzione di regressione alla fine del campione, inducendo, così, inevitabilmente a cattive previsioni.

Esistono due metodi principali per individuare le rotture appena descritte. La prima strategia cerca rotture potenziali secondo la prospettiva della verifica di ipotesi, ed implica la verifica della presenza di cambiamenti nei coefficienti di regressione utilizzando le statistiche F . La seconda strategia cerca tali rotture secondo la prospettiva delle previsioni: si assume che il campione termini prima di quando effettivamente accada e si valutano le previsioni che si sareb-

bero fatte in questo caso. Le rotture strutturali possono quindi essere individuate grazie al fatto che la *performance* delle previsioni è sostanzialmente peggiore di quanto ci si aspettava.

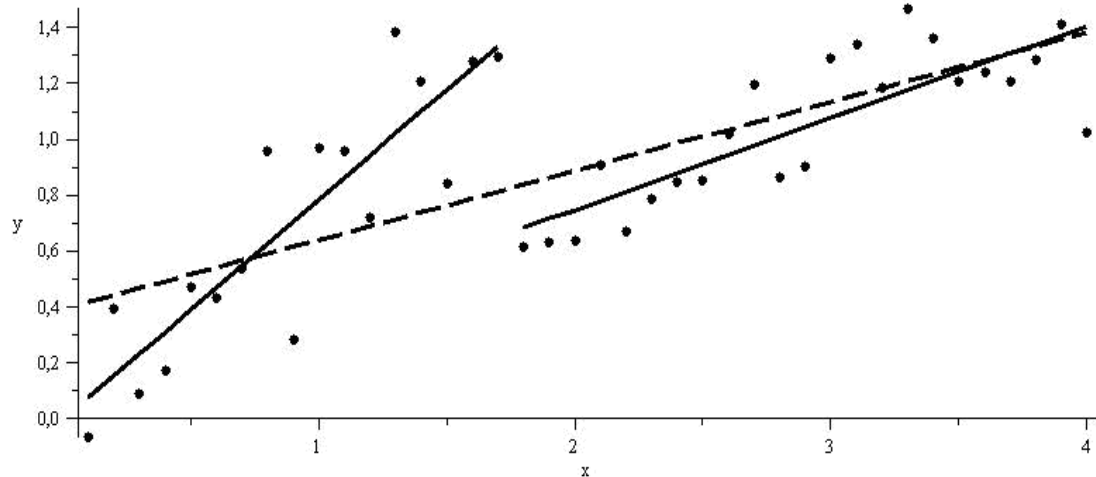


Figura 2.2: regressione lineare con break strutturale
(fonte: http://en.wikipedia.org/wiki/Chow_test).

Nella prossima sezione ci occuperemo della prima strategia menzionata, che risulterà poi utile nel terzo capitolo per considerare un'applicazione pratica dei metodi bootstrap passati in rassegna precedentemente.

2.2 Test per la verifica dell'ipotesi di *break* strutturale

Un modo per individuare le rotture strutturali è quello di verificare la presenza di cambiamenti netti nei coefficienti di regressione. Il metodo per fare ciò dipende dalla conoscenza o meno della possibile data di rottura (in inglese *break date*).

2.2.1 Test per la verifica di un *break* strutturale ad una data nota

In alcune applicazioni è possibile sospettare che vi sia una rottura ad una certa data. Ad esempio, se si studiano le relazioni di commercio internazionale utilizzando i dati degli anni Settanta, si potrebbe ipotizzare una rottura nella funzione di regressione temporale d'interesse nel 1972, quando il sistema di Bretton Woods di tassi di cambio fissi fu abbandonato in favore di tassi di cambio fluttuanti (vedi Figura 2.3).

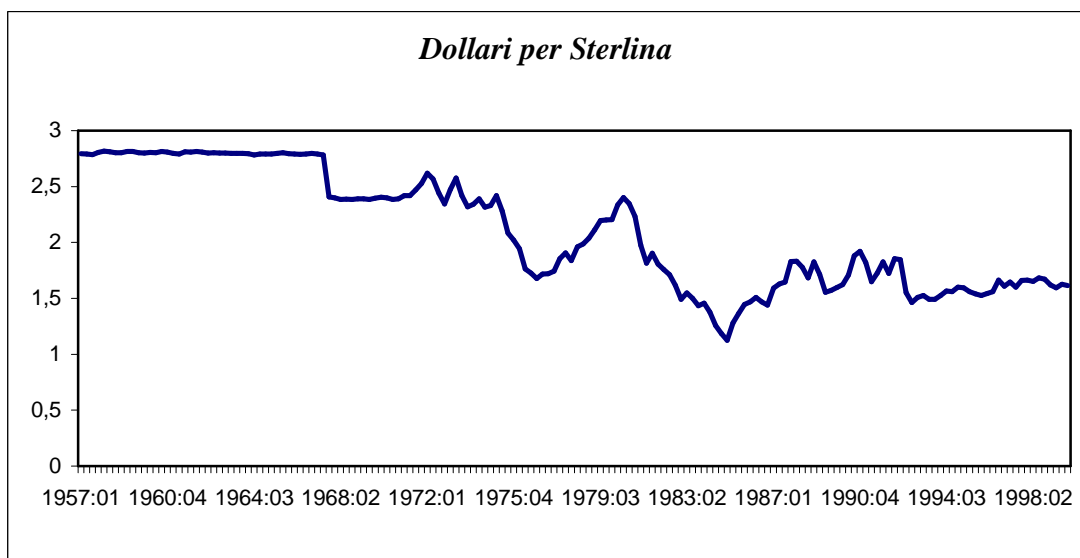


Figura 2.3: serie storica trimestrale del tasso di cambio dollaro/sterlina dal 1957 al 1999
(fonte dati: http://wps.aw.com/aw_stock_ie_2/50/13016/3332253.cw/index.html).

Se la data dell'ipotetica rottura nei coefficienti è nota, allora l'ipotesi nulla che non ci sia rottura può essere verificata utilizzando una regressione temporale. Ad esempio, si consideri per semplicità un modello autoregressivo ed a ritardi distribuiti ADL(1,1), in cui c'è un'intercetta, un singolo ritardo di Y_t , e un singolo ritardo di X_t . Sia τ l'ipotetica data della rottura strutturale e $D_t(\tau)$ una variabile binaria pari a zero prima della data di rottura e a uno dopo, ovvero $D_t(\tau) = 0$ se $t < \tau$ e $D_t(\tau) = 1$ se $t \geq \tau$. Allora la regressione che include la variabile *dummy* della rottura e tutti i termini d'interazione è:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \delta_1 X_{t-1} + \gamma_0 D_t(\tau) + \gamma_1 [D_t(\tau) \times Y_{t-1}] + \gamma_2 [D_t(\tau) \times X_{t-1}] + u_t,$$

dove $u_t \sim \text{i.i.d.}$, con media zero e varianza $\sigma^2 < \infty$.

Se non c'è una rottura strutturale, la funzione di regressione è la stessa in entrambe le porzioni del campione e, quindi, i termini che contengono l'indicatore binario per la rottura, $D_t(\tau)$, non entrano nella specificazione del modello dinamico preso in esame. Cioè, sotto l'ipotesi nulla di assenza di *break* strutturali, $\gamma_0 = \gamma_1 = \gamma_2 = 0$. Sotto l'ipotesi alternativa che ci sia un cambiamento di regime, la funzione di regressione è diversa prima e dopo la data τ della rottura, nel qual caso almeno uno dei γ è non nullo. Pertanto la presenza di una rottura strutturale può essere testata utilizzando la statistica F che verifica l'ipotesi di nullità congiunta dei coefficienti, $\gamma_0 = \gamma_1 = \gamma_2 = 0$, contro l'ipotesi che almeno uno dei γ sia non nullo. Questo test viene spesso chiamato *test di Chow* per una rottura strutturale ad una data nota, e deve il suo nome allo studioso Gregory Chow (1960). Se ci sono predittori multipli o più ritardi, questo test può essere esteso costruendo ulteriori termini d'interazione tra la variabile *dummy* $D_t(\tau)$ e i regressori aggiuntivi, verificando quindi l'ipotesi che tutti i coefficienti dei termini che coinvolgono $D_t(\tau)$ siano nulli. Inoltre, è possibile modificare questo approccio in modo tale da testare la presenza di un *break* strutturale in un particolare sottoinsieme dei coefficienti, includendo le interazioni con l'indicatore binario solo per tale sottoinsieme di regressori d'interesse.

Bisogna tuttavia precisare che, spesso, la data della rottura strutturale non è nota, anche se si pensa il contrario, e pure nel caso in cui si effettui una stima preliminare di tale data sulla base dei dati a disposizione, rimane comunque più appropriato adottare l'approccio seguente, utilizzato per verificare la presenza di un cambiamento di regime ad una data sconosciuta (Stock e Watson, 2005).

2.2.2 Test per la verifica di un *break* strutturale ad una data ignota

Spesso la data di una possibile rottura è ignota o è conosciuta solo entro un certo intervallo di tempo. Si supponga, ad esempio, che la rottura sia avvenuta in corrispondenza di una data compresa tra τ_0 e τ_1 . Il test di Chow può in questo caso essere modificato verificando la presenza di una rottura per tutte le possibili date τ tra τ_0 e τ_1 , e poi utilizzando la più grande delle statistiche F ottenute per testare la presenza di un *break* strutturale ad una data ignota. Questo test di Chow modificato è chiamato in vari modi, *statistica del rapporto delle verosimiglianze* o *statistica QLR* (acronimo dall'inglese *Quandt Likelihood Ratio*), *test di Quandt* (dal nome dello studioso Richard Quandt che per primo lo propose nel 1960), oppure, più raramente, *statistica sup di Wald* (in inglese *sup-Wald statistic*). Formalmente, la statistica QLR è pari a:

$$QLR = \max [F(\tau_0), F(\tau_0 + 1), \dots, F(\tau_1)].$$

Poiché il test di Quandt è la più grande tra molte statistiche F , la sua distribuzione sotto l'ipotesi nulla non è la stessa di una statistica F individuale. I valori critici per la statistica QLR devono quindi essere ottenuti da una distribuzione speciale. Come la statistica F , questa distribuzione dipende dal numero di restrizioni che vengono testate, q , cioè dal numero di coefficienti (intercetta inclusa) che possono essere soggetti a rottura, o cambiamento, sotto l'ipotesi alternativa. La distribuzione del test di Quandt sotto l'ipotesi di stabilità dei coefficienti dipende anche da τ_0/T e τ_1/T , cioè dai punti estremi τ_0 e τ_1 del sottocampione per il quale le statistiche F sono calcolate, espressi come frazione della numerosità campionaria totale.

Affinché l'approssimazione per grandi campioni della distribuzione della statistica QLR sia buona, i punti estremi del sottocampione, τ_0 e τ_1 , non possono essere troppo vicini agli estremi dell'intero campione a disposizione. Per questa ragione, nella pratica, il test di Quandt è calcolato su un intervallo troncato, o un sottoinsieme, del campione. Una scelta comune è quella di utilizzare

un troncamento del 15%, fissando cioè $\tau_0 = 0,15T$ e $\tau_1 = 0,85T$ (arrotondati all'intero più vicino). Con un troncamento al 15%, la statistica F è calcolata per date di rottura corrispondenti al 70% centrale del campione di dati.

I valori critici della distribuzione asintotica della statistica QLR derivati da Andrews (1993), e calcolati con un troncamento del 15%, sono mostrati nella Tabella 2.1. Il confronto di questi valori con quelli della distribuzione F di Snedecor rivela che i valori critici per il test di Quandt sono più ampi. Questo ha senso, perché riflette il fatto che la statistica QLR considera la più grande delle statistiche F individuali. Esaminando le statistiche F in prossimità di molte possibili date di rottura, la statistica QLR ha molte opportunità di rifiutare, determinando valori critici della statistica QLR che sono maggiori di quelli delle statistiche F individuali.

Tabella 2.1: valori critici del test di Quandt per α pari a 10%, 5% ed 1% con un troncamento del 15% (fonte: Stock e Watson, 2005).

<i>Numero di restrizioni (q)</i>	10%	5%	1%
1	7,12	8,68	12,16
2	5,00	5,86	7,78
3	4,09	4,71	6,02
4	3,59	4,09	5,12
5	3,26	3,66	4,53
6	3,02	3,37	4,12
7	2,84	3,15	3,82
8	2,69	2,98	3,57
9	2,58	2,84	3,38
10	2,48	2,71	3,23
11	2,40	2,62	3,09
12	2,33	2,54	2,97
13	2,27	2,46	2,87
14	2,21	2,40	2,78
15	2,16	2,34	2,71
16	2,12	2,29	2,64
17	2,08	2,25	2,58
18	2,05	2,20	2,53
19	2,01	2,17	2,48
20	1,99	2,13	2,43

Se nella serie storica è presente una rottura strutturale netta ad una certa data all'interno dell'intervallo temporale in esame, allora la statistica QLR rifiuterà l'ipotesi nulla di assenza di un *break* strutturale con alta probabilità per grandi campioni, quindi la potenza del test è soddisfacente per numerosità campionarie elevate. Inoltre, la data in cui la statistica F che compone il test di Quandt raggiunge il suo massimo, $\hat{\tau}$, è una stima della data τ . Questa stima è buona, nel senso che, sotto certe condizioni tecniche, $\hat{\tau}/T \xrightarrow{p} \tau/T$, cioè la frazione dell'intervallo campionario dopo la quale si ha la rottura è stimata consistentemente (Stock e Watson, 2005).

La statistica QLR rifiuta l'ipotesi nulla con alta probabilità per grandi campioni anche quando ci sono multiple rotture strutturali nette oppure quando il cambiamento di regime prende la forma di una lenta evoluzione della funzione di regressione. Ciò significa che il test di Quandt può individuare forme di instabilità diverse da una singola rottura netta. Come risultato, se la statistica QLR rifiuta l'ipotesi nulla di assenza di cambiamento strutturale nella serie storica oggetto di analisi, allora ci possono essere una singola rottura netta, molteplici rotture nette, oppure una graduale evoluzione della funzione di regressione temporale.

Nel prossimo capitolo, considereremo il test di Quandt per dare un semplice esempio dell'applicazione dei diversi metodi bootstrap per serie storiche precedentemente descritti. In particolare, li utilizzeremo per calcolare una stima dell'ampiezza effettiva del test.

Capitolo 3

APPLICAZIONE DEL BOOTSTRAP

3.1 Introduzione

Come già evidenziato in precedenza, il bootstrap consente di stimare consistentemente media e varianza di un certo stimatore, costruire intervalli di confidenza, e calcolare *p-values* effettivi di test quando, in particolare, la distribuzione esatta sotto l'ipotesi nulla della statistica test d'interesse non è nota e/o la sua distribuzione asintotica non porta a risultati soddisfacenti (per esempio in presenza di campioni finiti). Infatti, il bootstrap si rivela particolarmente utile quando si tratta di approssimare la funzione di distribuzione campionaria di uno stimatore o di una statistica d'interesse.

In questa tesi, si è scelto di fornire come esempio dell'applicazione dei metodi bootstrap per serie storiche trattati in precedenza, la stima della “vera” probabilità dell'errore di I tipo della statistica test QLR (Quandt, 1960), utilizzata per verificare la presenza di rotture strutturali ad una data ignota. Come già spiegato nel capitolo precedente, la distribuzione di questa statistica è differente dalla distribuzione F di Snedecor, e in particolare i valori critici della sua distribuzione asintotica, comunemente utilizzati e riportati in Tabella 2.1, sono più ampi rispetto a quelli della F di Snedecor, e dipendono sia dal numero di restrizioni q da verificare sia dai punti estremi τ_0 e τ_1 come frazioni della numerosità campionaria T .

Abbiamo quindi considerato, come DGP, processi stocastici autoregressivi di ordine uno e due. Nel caso di modelli AR(1), la specificazione del test di Quandt si basa sulle singole statistiche F il cui modello allargato sotto l'ipotesi alternativa è:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \gamma_0 D_t(\tau) + \gamma_1 [D_t(\tau) \times Y_{t-1}] + \varepsilon_t,$$

dove $D_t(\tau)$ è una variabile *dummy* definita come nel capitolo precedente. Quindi, sotto l'ipotesi nulla della stabilità dei coefficienti, $\gamma_0 = \gamma_1 = 0$. Il numero di restrizioni q per ogni singola statistica F è pari a due e, utilizzando un troncamento del campione al 15%, il valore critico della statistica QLR al livello di significatività del 5%, con cui confrontare il valore ottenuto, è 5.86 (vedi Tabella 2.1 nel capitolo precedente).

Nel caso di modelli AR(2) invece, la statistica QLR costituisce il valore massimo delle statistiche F calcolate sulla base della specificazione, sotto l'ipotesi alternativa, del modello completo che prende questa forma:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \gamma_0 D_t(\tau) + \gamma_1 [D_t(\tau) \times Y_{t-1}] + \gamma_2 [D_t(\tau) \times Y_{t-2}] + \varepsilon_t,$$

dove $D_t(\tau)$ è ancora una variabile *dummy* definita come nel secondo capitolo. Pertanto, sotto l'ipotesi nulla di assenza di cambiamenti strutturali, $\gamma_0 = \gamma_1 = \gamma_2 = 0$. Il numero di restrizioni da testare è perciò $q=3$, e utilizzando un troncamento del 15%, il valore critico corrispondente del test di Quandt, al livello del 5%, è pari a 4.71 (vedi Tabella 2.1).

3.2 Esperimenti di simulazione Monte Carlo

Per valutare se la distribuzione asintotica della statistica QLR fornisce una buona approssimazione della vera distribuzione sconosciuta del test sotto l'ipotesi nulla, abbiamo determinato una stima dell'effettiva probabilità dell'errore di I tipo (cioè la probabilità di rigettare H_0 , posto H_0 vera) attraverso degli esperimenti di simulazione di tipo Monte Carlo. In particolare, abbiamo effettuato queste simulazioni per differenti valori dei parametri del processo generatore dei dati e per differenti lunghezze delle serie storiche generate.

Abbiamo posto il numero di simulazioni s pari a 1000, e nel caso di modelli AR(1), abbiamo fatto variare il parametro autoregressivo ϕ_1 entro un insieme di sei valori possibili (0.3, 0.6, 0.9, -0.3, -0.6, -0.9), utilizzando sia una lunghez-

za T delle serie storiche generate pari a 50 sia pari a 100. Per determinare il “vero” errore di I tipo del test di Quandt (quando confrontato coi valori critici della sua distribuzione asintotica), grazie al supporto del noto software R per analisi statistiche (*www.r-project.org*) abbiamo simulato, per ogni diverso valore di ϕ_1 e di T , 1000 serie storiche senza rotture strutturali (sotto H_0), e ricavato per ognuna di esse la statistica QLR per testare la stabilità dei coefficienti. Abbiamo poi calcolato la proporzione di statistiche test maggiori del valore critico al livello di significatività nominale del 5% riportato in Tabella 2.1, 5.86, ottenendo così una stima della probabilità effettiva (o empirica) dell’errore di I tipo del test. I risultati ottenuti sono mostrati in Tabella 3.1.

Tabella 3.1: risultati di esperimenti Monte Carlo per ricavare la size effettiva del test di Quandt su modelli AR(1) con numero di simulazioni pari a 1000.

ϕ_1	0.3	0.6	0.9	-0.3	-0.6	-0.9
$T=50$	5.7%	9.8%	22%	4%	3.5%	4.1%
$T=100$	5.9%	8.2%	20.3%	3.2%	3.2%	4.2%

Dai valori riportati in Tabella 3.1, si può constatare che la *size* effettiva stimata del test di Quandt (ovvero il livello di significatività effettivo stimato) per valori di ϕ_1 pari a 0.6 e a 0.9 risulta essere particolarmente distorta rispetto al valore nominale, o teorico, del 5% (per la precisione, ciò si verifica in maniera molto più netta per $\phi_1=0.9$), e questa distorsione è leggermente maggiore per serie storiche di numerosità campionaria più limitata ($T=50$).

Il codice R utilizzato per ottenere i precedenti risultati, simulando, ad esempio, 1000 serie storiche con parametro autoregressivo pari a 0.6 e composte da 100 osservazioni è riportato in Appendice.

Come si può vedere dalle righe di codice, nella funzione QLR che calcola e restituisce il test di Quandt con troncamento del 15% per una data serie storica, si utilizza la libreria di R `dynlm` per modelli dinamici. Dopo aver calcolato la statistica test per le 1000 serie storiche con parametro autoregressivo pari a quello fissato e simulate sotto l'ipotesi nulla con la funzione `arima.sim`, si può facilmente ottenere la frazione di valori osservati della statistica QLR maggiori del valore critico al livello nominale del 5%, 5.86 (cv), la quale approssima in modo consistente la probabilità di rifiutare H_0 , condizionatamente al fatto che H_0 sia vera, dove H_0 è l'ipotesi di assenza di rotture strutturali nella serie storica oggetto di analisi.

Per i modelli AR(2), analogamente al caso dei modelli autoregressivi di ordine uno, abbiamo effettuato esperimenti di simulazione di tipo Monte Carlo per i tre differenti valori delle coppie di parametri (-1, -0.2), (-0.3, -0.9) e (0.5, 0.4), sia per numerosità campionarie pari a 50, sia per numerosità pari a 100. Abbiamo scelto proprio queste tre coppie perché ben distribuite all'interno del *Triangolo di stazionarietà* dei processi AR(2), ma non ci aspettiamo risultati sensibilmente diversi per altre coppie di coefficienti presenti all'interno della suddetta regione triangolare.

Ponendo sempre il numero di simulazioni $s=1000$, abbiamo ottenuto i risultati mostrati in Tabella 3.2., che danno evidenza di una particolare distorsione nella effettiva probabilità dell'errore di I tipo, rispetto al valore teorico del 5%, per la coppia di parametri $\phi_1=0.5$ e $\phi_2=0.4$, distorsione solo lievemente superiore per la numerosità campionaria più limitata ($T=50$).

Tabella 3.2: risultati di esperimenti Monte Carlo per ricavare la size effettiva del test di Quandt su modelli AR(2) con numero di simulazioni pari a 1000.

	$\phi=(-1, -0.2)$	$\phi=(-0.3, -0.9)$	$\phi=(0.5, 0.4)$
T=50	4%	5.1%	18.9%
T=100	4.2%	4.7%	18.3%

Anche in questo caso, le righe di codice R utilizzate per ricavare questi risultati, simulando, ad esempio, 1000 serie storiche di lunghezza pari a 50 con parametri autoregressivi posti uguali a $(-1, -0.2)$, sono riportate in Appendice. Come per il caso precedente, nella funzione QLR che calcola il test di Quandt con troncamento al 15% per la serie storica data in input, si utilizza la libreria di R `dynlm` per modelli dinamici. Vediamo che nel caso AR(2) si deve tener conto della presenza del secondo ritardo della variabile dipendente tra i regressori, sia nella formulazione del modello ridotto sotto l'ipotesi nulla (`mod_0`) sia in quella del modello allargato sotto l'ipotesi alternativa (`mod_1`). Dopo aver attribuito i valori fissati ai parametri ϕ e alla numerosità campionaria T , si simulano 1000 serie storiche da processi stocastici autoregressivi di ordine due senza *break* strutturali (quindi sotto l'ipotesi nulla di assenza di cambiamenti di regime nella serie storica presa in esame), con parametri pari ai valori fissati, e si ricavano le mille statistiche di Quandt applicando ad ogni serie generata la funzione QLR. A questo punto è possibile calcolare la proporzione di test di Quandt osservati maggiori del corrispondente valore critico della distribuzione asintotica, 4.71 (`cv`), al livello di significatività nominale del 5% e definito in Tabella 2.1 (ricordiamo che nel caso AR(2) si verifica la nullità di tre coefficienti con le singole statistiche F che compongono il test). Così facendo, stimiamo la “vera” probabilità dell'errore di I tipo della statistica QLR per modelli AR(2).

3.3 Applicazione del bootstrap

Si è già discusso più volte in questa tesi dei vantaggi apportati dall'applicazione della procedura bootstrap quando si è interessati a stimare in modo consistente la funzione di distribuzione campionaria di una statistica test d'interesse. Sulla base dei risultati degli esperimenti Monte Carlo riportati nella sezione precedente, abbiamo deciso di applicare il bootstrap a serie storiche simulate per verificare se la stima dell'ampiezza effettiva del test di Quandt così ricavata risultasse migliorata, ovvero più vicina al valore nominale del 5%, rispetto a

quella ottenuta con gli esperimenti Monte Carlo per determinati valori dei coefficienti autoregressivi. Purtroppo, a causa del grosso carico computazionale richiesto dalla procedura bootstrap, non è stato possibile effettuare un numero di simulazioni s e un numero di replicazioni bootstrap B tali da analizzare in modo totalmente soddisfacente la sua *performance*. Tuttavia, riteniamo che i risultati conseguiti, anche se con una quantità di simulazioni e di replicazioni inferiore a quella ottimale, possano essere utili per valutare, almeno indicativamente, il grado di precisione raggiunto dai tre metodi bootstrap presi in esame nel primo capitolo.

Nella sezione precedente si è evidenziato come, nel caso di modelli autoregressivi di ordine uno, la probabilità “vera”, stimata con gli esperimenti Monte Carlo, dell’errore di I tipo del test di Quandt risultasse particolarmente distorta rispetto al suo valore nominale per valori del coefficiente AR pari a 0.6 e 0.9 (vedi Tabella 3.1). Si è quindi deciso di verificare se, applicando il bootstrap a serie storiche simulate con questi valori del coefficiente autoregressivo, si ottenesse un maggior rispetto dell’ampiezza teorica, ovvero se la stima della dimensione empirica del test risultasse più vicina al livello nominale del 5% o meno.

Per ognuno dei due valori di ϕ_1 , 0.6 e 0.9, sono state simulate $s=200$ serie storiche sotto l’ipotesi nulla di stabilità strutturale, di lunghezza $T=50$ o $T=100$. A ciascuna serie generata sono stati poi applicati i tre principali metodi bootstrap considerati in questa tesi (il *model-based resampling*, il *block bootstrap* e il *sieve bootstrap*) per ricavare la distribuzione empirica della statistica test QLR e calcolarne il 95-esimo percentile, in modo tale da poter stimare consistentemente, per ogni diverso tipo di bootstrap, l’effettiva probabilità di rigettare l’ipotesi nulla, quando l’ipotesi nulla è invece vera.

Le righe di codice R utilizzate per analizzare il caso, ad esempio, di serie storiche simulate con parametro autoregressivo ϕ_1 pari a 0.9 e dimensione campionaria $T=100$, sono riportate in Appendice. Da queste vediamo che si inizializzano, ponendoli pari a zero, tre contatori relativi ai tre metodi bootstrap utilizzati, e si attribuiscono i valori prefissati al parametro autoregressivo e alla

numerosità campionaria T . La funzione `Qmboot` riceve in input una serie storica y e un quantile p -esimo ($p=0.95$), quindi genera $B=200$ serie storiche bootstrap col metodo *model-based resampling*, la cui procedura è spiegata in dettaglio nel primo capitolo, e calcola per ognuna di esse la statistica test di Quandt attraverso la funzione `QLR`. Questi duecento valori del test costituiscono l'approssimazione bootstrap della distribuzione campionaria della statistica `QLR`, per cui è possibile calcolarne il 95-esimo percentile, che costituisce anche l'output della funzione `Qmboot`. Analogamente, la funzione `Qsboot` applica alla serie data in input il metodo *sieve bootstrap* per generare $B=200$ serie storiche necessarie ad ottenere un'approssimazione della distribuzione empirica del test di Quandt, per poi ricavarne il quantile 0.95. Per effettuare il bootstrap a blocchi, e precisamente il *moving-block bootstrap*, si utilizza la funzione `tsboot`, contenuta nella libreria `boot`, che riceve in input una serie storica y , la funzione da applicare alle serie bootstrap generate, il numero $R=200$ delle repliche bootstrap da generare, la lunghezza dei blocchi da ricampionare e il metodo del bootstrap a blocchi da utilizzare (si noti che, per impostazione di default, la funzione `tsboot` utilizza blocchi sovrapposti). In questo caso, poniamo la funzione `QLR` come quella da applicare alle serie storiche bootstrap, una lunghezza dei blocchi pari a 5 (adottando la regola del pollice seguita da Bühlmann, 2002, che utilizza una dimensione dei blocchi pari a $T^{1/3}$ approssimato all'intero più vicino, quindi $100^{1/3} \approx 5$), e il ricampionamento di blocchi di lunghezza fissa $l=5$ come tipo di simulazione richiesta per generare le repliche bootstrap (`sim="fixed"`). Ancora una volta si determina il 95-esimo percentile della distribuzione bootstrap della statistica `QLR`, ricavata dai duecento valori del test relativi alle serie storiche bootstrap generate. Quindi, per ognuna delle $s=200$ serie storiche, simulate sotto l'ipotesi nulla di assenza di rotture strutturali, si calcola la statistica `QLR` osservata e ne si confronta il valore coi tre quantili 0.95 delle distribuzioni campionarie del test ottenute mediante l'applicazione dei tre differenti metodi bootstrap ad ogni serie simulata. Se questo valore osservato del test di Quandt risulta maggiore del 95-esimo percentile calcolato con un certo metodo bootstrap, il contatore relativo a quel

metodo viene incrementato di un'unità. Alla fine, per verificare se, utilizzando i valori critici determinati grazie all'impiego del bootstrap, la probabilità effettiva stimata dell'errore di I tipo del test QLR risulti o meno più vicina all'ampiezza teorica del 5% rispetto a quella stimata utilizzando i valori critici della distribuzione asintotica (vedi esperimenti Monte Carlo riportati nella sezione precedente), si calcolano i rapporti tra i valori dei contatori e il numero totale s di serie storiche simulate, così da ottenere una stima della *size* empirica del test per ogni metodo bootstrap impiegato. I risultati conseguiti grazie al codice R appena analizzato, per $\phi_1=0.6$ e $\phi_1=0.9$, con $T=50$ e $T=100$, sono riportati, rispettivamente, nelle Tabelle 3.3 e 3.4.

Tabella 3.3: risultati dell'applicazione del bootstrap per ricavare la size effettiva del test di Quandt a modelli AR(1) con $\phi_1=0.6$, $s=200$ e $B=200$.

Bootstrap	<i>model-based</i>	<i>block</i>	<i>sieve</i>
$T=50$	5%	1.6% ($l=4$)	5.5%
$T=100$	5.5%	2% ($l=5$)	3.5%

Tabella 3.4: risultati dell'applicazione del bootstrap per ricavare la size effettiva del test di Quandt a modelli AR(1) con $\phi_1=0.9$, $s=200$ e $B=200$.

Bootstrap	<i>model-based</i>	<i>block</i>	<i>sieve</i>
$T=50$	8%	2.5% ($l=4$)	7%
$T=100$	8%	2% ($l=5$)	8%

I risultati danno evidenza di un notevole miglioramento nella stima della *size* effettiva del test di Quandt ottenuta applicando il bootstrap, rispetto a quella

determinata con gli esperimenti di simulazione Monte Carlo (vedi Tabella 3.1): questi valori sono infatti molto più vicini dei precedenti al livello nominale del 5%. In particolare, le stime ricavate avvalendosi di serie storiche con valore del parametro autoregressivo pari a 0.6 rispettano l'ampiezza nominale più delle altre, ed in effetti, per questo valore di ϕ , anche la distorsione iniziale delle *size* stimate col metodo Monte Carlo era decisamente minore. Notiamo, invece, che non vi sono differenze evidenti, in termini di stime ottenute, al variare della numerosità campionaria per ciascuno dei valori di ϕ . Inoltre, osserviamo che il *moving-block bootstrap*, al contrario degli altri due metodi di ricampionamento, tende a restituire stime della *size* effettiva che sono inferiori al livello nominale del 5%. Questo ci porta a concludere che, quando confrontato con valori critici ricavati mediante il bootstrap a blocchi, il test di Quandt diventa un test conservativo, ovvero ha la tendenza ad accettare l'ipotesi nulla più frequentemente di quanto atteso, quando l'ipotesi nulla è vera. Bisogna poi considerare che, specialmente per $\phi=0.9$, la discrepanza tra probabilità empirica e probabilità teorica dell'errore di I tipo del test non è interamente corretta dall'impiego dei metodi bootstrap. Nel nostro caso specifico, parte della responsabilità di questa distorsione residua potrebbe essere attribuita al modesto numero di simulazioni s e di repliche B che si sono potute effettuare a causa dell'importante carico computazionale richiesto dalla procedura bootstrap. È infatti possibile che, con un numero di simulazioni adeguato ($s \geq 500$) ed una maggior quantità di repliche bootstrap B , i valori ottenuti dalle stime dei *p-values* effettivi del test si sarebbero avvicinati ulteriormente al livello di significatività nominale del 5%.

In modo analogo a quanto descritto per i processi AR(1), abbiamo cercato di investigare gli effetti provocati dall'applicazione del bootstrap a serie storiche AR(2) sulle stime della *size* empirica del test di Quandt. Facendo riferimento ai risultati degli esperimenti Monte Carlo relativi alla distribuzione asintotica del test e mostrati nella precedente sezione, abbiamo preso in considerazione la coppia di parametri autoregressivi per cui la probabilità "vera" stimata dell'errore di I tipo del test risultava essere particolarmente distante dal livello

nominale del 5%, quindi la coppia di coefficienti $\phi_1=0.5$ e $\phi_2=0.4$ (vedi Tabella 3.2). Purtroppo, a causa di difficoltà legate al costo computazionale e ai tempi di simulazione richiesti dalla procedura bootstrap per modelli AR(2), abbiamo dovuto concentrarci soltanto su due metodi, ovvero il *model-based resampling* e il *sieve bootstrap*. Come sempre, in Appendice riportiamo il codice R usato (nel caso di $T=50$) per ottenere i risultati. Notiamo che le righe di codice utilizzate per applicare il bootstrap a modelli autoregressivi di ordine due sono molto simili a quelle utilizzate per modelli AR(1), ed infatti la procedura seguita è analoga a quella già precedentemente analizzata. Le principali differenze riguardano la funzione QLR, che deve tenere conto della presenza del secondo ritardo della variabile dipendente tra i regressori, nei modelli dinamici creati sotto H_0 e sotto H_1 , la presenza di due parametri autoregressivi che devono essere stimati all'interno della funzione `Qmbot` che effettua il *model-based resampling* bootstrap, e naturalmente la simulazione, sotto l'ipotesi nulla di stabilità strutturale, di $s=200$ serie storiche realizzazioni finite di processi stocastici AR(2), con valori dei parametri opportunamente fissati pari a quelli d'interesse.

Le stime della “vera” probabilità dell'errore di I tipo del test di Quandt, ottenute grazie all'impiego della metodologia bootstrap per determinare i valori critici costituiti dai 95-esimi percentili delle distribuzioni empiriche del test (approssimate utilizzando il *parametric* e il *sieve bootstrap*), sono riportate in Tabella 3.5.

Tabella 3.5: risultati dell'applicazione del bootstrap per ricavare la size effettiva del test di Quandt a modelli AR(2) con $\phi=(\mathbf{0.5}, \mathbf{0.4})$, $s=200$ e $B=200$.

Bootstrap	<i>model-based</i>	<i>sieve</i>
$T=50$	9.5%	8.5%
$T=100$	8%	6.5%

Dall'analisi della Tabella 3.5 rileviamo che, per la numerosità campionaria più limitata ($T=50$), la metodologia bootstrap migliora considerevolmente la frequenza di rifiuto sotto H_0 , e questo miglioramento diventa anche maggiore quando la lunghezza delle serie storiche simulate è pari a 100. Infatti, rispetto alla stima ottenuta attraverso gli esperimenti Monte Carlo basati sulla distribuzione asintotica del test, 18.9% (vedi Tabella 3.2), l'avvicinamento al valore nominale è notevole e, considerando pure il numero decisamente limitato di simulazioni effettuate e la quantità inferiore a quella ottimale di replicazioni bootstrap utilizzate, piuttosto soddisfacente.

3.4 Conclusioni

Con questo semplice esempio di applicazione del bootstrap a serie storiche simulate, si è voluto mostrare come l'accuratezza nelle stime raggiunta con questo metodo sia significativamente superiore a quella raggiunta per mezzo delle approssimazioni asintotiche del primo ordine. Queste ultime, infatti, sono spesso imprecise ed ingannevoli quando applicate a dati di serie storiche e a campioni di dimensioni finite comunemente a disposizione dei ricercatori.

In questo capitolo abbiamo cercato di stimare in modo consistente la dimensione effettiva del test di Quandt. Per fare ciò, bisogna conoscere la distribuzione del test sotto l'ipotesi nulla: tradizionalmente, viene utilizzata quella asintotica, valida solo se la dimensione campionaria è infinitamente elevata. Nel nostro caso, mediante l'effettuazione di esperimenti di simulazione di tipo Monte Carlo, abbiamo verificato che la *size* empirica del test è alquanto differente da quella teorica quando il valore osservato della statistica viene confrontato con gli usuali valori critici della distribuzione asintotica. Dai risultati si vede che, in particolare, la probabilità di commettere l'errore di I tipo è spesso maggiore rispetto alla probabilità attesa, e questo può portare a risultati fuorvianti nella verifica di ipotesi. Pertanto è possibile concludere che la distribuzione asintotica del test, almeno quando questo viene applicato a serie storiche AR(1) e AR(2) con determinati valori dei parametri autoregressivi e di lunghez-

za limitata, si discosta notevolmente dalla reale ignota distribuzione del test di Quandt sotto l'ipotesi nulla. Le tecniche bootstrap costituiscono un modo alternativo di ottenere le distribuzioni (e quindi i valori critici) con cui i valori osservati delle statistiche test devono essere confrontati. Abbiamo quindi utilizzato i tre metodi bootstrap analizzati in questa tesi per ricavare delle approssimazioni più accurate della vera distribuzione sconosciuta del test di Quandt sotto l'ipotesi nulla di assenza di *break* strutturali. Le distribuzioni bootstrap della statistica QLR sono state ottenute generando, a partire da una data serie simulata, un gran numero di pseudo-serie bootstrap, e calcolando per ognuna di esse la statistica d'interesse. Abbiamo, quindi, nuovamente stimato la *size* effettiva del test per ogni metodo di ricampionamento utilizzato, confrontando il valore osservato del test per ogni serie simulata coi valori critici ricavati dalle distribuzioni bootstrap della statistica QLR. Le stime così determinate (seppur con un numero limitato di simulazioni e di repliche bootstrap) sono risultate essere molto più vicine al valore dell'ampiezza teorica rispetto a quelle basate sulla distribuzione asintotica del test, suggerendo che il bootstrap ha effettivamente approssimato in modo più preciso la distribuzione del test di Quandt sotto l'ipotesi nulla.

Il bootstrap riesce quindi a correggere la distorsione nella stima della dimensione empirica del test di Quandt, portando ad un maggior rispetto della ampiezza teorica. Più precisamente, come dimostrano i risultati, i metodi del *model-based resampling* e del *sieve bootstrap* tendono a sovrastimare la *size* nominale, mentre il *moving-block bootstrap* la sottostima. In quest'ultimo caso, il test si rivela conservativo, perché, quando l'ipotesi nulla è vera, il test la rifiuta meno spesso di quanto ci si attende. Comunque, tutti e tre i metodi migliorano notevolmente la frequenza di rifiuto del test sotto H_0 , rispetto a quando si utilizzano i valori critici relativi alla distribuzione asintotica. Questo non ci sorprende, visto che nel primo capitolo già abbiamo evidenziato come i risultati raggiunti da queste tecniche siano soddisfacenti quando si verificano determinate condizioni. In particolare, per il *parametric bootstrap*, quando la specificazione del modello è corretta, per il *block bootstrap*, quando la serie storica osservata è realizzazione di un processo stocastico con determinate

proprietà, quali la stazionarietà e la dipendenza temporale di breve periodo, e per il *sieve bootstrap*, quando il processo generatore dei dati rispetta le assunzioni a priori necessarie, tra cui la linearità e l'invertibilità. Nel nostro caso, le serie storiche simulate erano realizzazioni di processi autoregressivi di ordine finito (uno o due) stazionari ed invertibili, con errori omoschedastici indipendenti ed identicamente distribuiti, quindi le condizioni suddette erano pienamente soddisfatte.

Appendice

In questa appendice vengono riportate le righe di codice R utilizzate per ottenere i risultati, riportati nel terzo capitolo, delle simulazioni Monte Carlo e delle applicazioni del bootstrap.

A.1 Codice R per gli esperimenti di simulazione Monte Carlo

Il codice R utilizzato per gli esperimenti Monte Carlo relativi ai modelli AR(1) (vedi sezione 3.2) ed effettuati simulando, ad esempio, 1000 serie storiche con parametro autoregressivo $\phi_1 = 0.6$ e composte da 100 osservazioni, è il seguente:

```
rm(list = ls())

library(dynlm)

QLR <- function(y) {
  tau_0 <- round(0.15*length(y))
  tau_1 <- round(0.85*length(y))
  Q <- NULL
  mod_0 <- dynlm(y~L(y))
  for (i in tau_0:tau_1) {
    n1 <- i
    n2 <- length(y)-i
    D <- ts(c(rep(0,n1), rep(1,n2)))
    mod_1 <- dynlm(y~L(y)+D+D*L(y))
    t_C <- anova(mod_0, mod_1)$F[2]
    Q[1+i-tau_0] <- t_C
  }
  max(Q)
}
```

```

s <- 1000
phi <- 0.6
T <- 100

cv <- 5.86
alpha <- 1
Q <- NULL

for (i in 1:s) {
  x <- arima.sim(list(order=c(1,0,0), ar=phi),n=T)
  y <- x+alpha
  Q[i] <- QLR(y)
}

mean(Q>cv)

```

Le righe di codice R descritte nella sezione 3.2 e utilizzate per effettuare gli esperimenti di tipo Monte Carlo relativi ai modelli autoregressivi di ordine 2, sono riportate di seguito. In particolare, in questo caso vengono simulate 1000 serie storiche di lunghezza $T=50$ e con parametri autoregressivi posti uguali a $(-1, -0.2)$:

```

rm(list = ls())

library(dynlm)

QLR <- function(y) {
  tau_0 <- round(0.15*length(y))
  tau_1 <- round(0.85*length(y))
  Q <- NULL
  mod_0 <- dynlm(y~L(y, 1:2))
  for (i in tau_0:tau_1) {
    n1 <- i

```

```

n2 <- length(y)-i
D <- ts(c(rep(0,n1), rep(1,n2)))
mod_1 <- dynlm(y~L(y, 1:2)+D+D*L(y, 1:2))
t_C <- anova(mod_0, mod_1)$F[2]
Q[1+i-tau_0] <- t_C
}
max(Q)
}

s <- 1000
phi_1 <- -1
phi_2 <- -0.2
T <- 50

cv <- 4.71
alpha <- 1
Q <- NULL

for (i in 1:s) {
  x <- arima.sim(list(order=c(2,0,0), ar=c(phi_1,
  phi_2)), n=T)
  y <- x+alpha
  Q[i] <- QLR(y)
}

mean(Q>cv)

```

A.2 Codice R per l'applicazione del bootstrap

I risultati riportati in sezione 3.3 e conseguiti grazie all'applicazione del bootstrap a serie storiche AR(1) con, ad esempio, parametro autoregressivo pari a 0.9 e dimensione campionaria $T=100$, sono stati ottenuti mediante il seguente codice R:

```

rm(list=ls())

library(dynlm)
library(boot)

QLR <- function(y) {
  tau_0 <- round(0.15*length(y))
  tau_1 <- round(0.85*length(y))
  Q <- NULL
  mod_0 <- dynlm(y~L(y))
  for (i in tau_0:tau_1) {
    n1 <- i
    n2 <- length(y)-i
    D <- c(rep(0,n1), rep(1,n2))
    mod_1 <- dynlm(y~L(y)+D+D*L(y))
    t_C <- anova(mod_0, mod_1)$F[2]
    Q[1+i-tau_0] <- t_C
  }
  max(Q)
}

contb <- 0
contm <- 0
conts <- 0

s <- 200
phi <- 0.9
T <- 100

alpha <- 1
q <- NULL
p <- 0.95

Qmboot <- function(y, p) {
  Q <- NULL

```



```

n <- length(y)
mod <- arima(y, order=c(1,0,0))
phi_hat <- mod$coef[1]
media_hat <- mod$coef[2]
residui <- na.omit(mod$res)
c.res <- residui-mean(residui)
B <- 200
cost <- (1-phi_hat)*media_hat
for (i in 1:B) {
  eb <- sample(c.res, n, replace=T)
  xb <- arima.sim(list(ar=phi_hat), mean=cost,
    innov=eb, n)
  Q[i] <- QLR(ts(xb))
}
quantile(Q, p)
}

```

```

Qsboot <- function(x, p) {
  Q <- NULL
  n <- length(x)
  y <- x-mean(x)
  fit <- ar(y)
  pstim <- fit$order
  phi.stim <- fit$ar
  residui <- na.omit(fit$res)
  res.c <- residui-mean(residui)
  B <- 200
  for (i in 1:B) {
    if (pstim == 0) {
      yb <- sample(res.c, n, replace=T)
      xb <- yb + mean(x)
    }
    if (pstim > 0) {
      yb <- NULL
      eb <- sample(res.c, n, replace=T)

```

```

        yb[1:pstim] <- mean(y)
        for (j in (pstim+1):n) {
            yb[j] <- (phi.stim)**(yb[(j-1):(j-
                pstim)]) + eb[j]
        }
        xb <- yb + mean(x)
    }
    Q[i] <- QLR(ts(xb))
}

quantile(Q, p)
}

for (i in 1:s) {
    x <- arima.sim(list(order=c(1,0,0), ar=rho), n=T)
    y <- x+alpha
    q <- tsboot(y, QLR, R=200, l=5, sim="fixed")$t
    Qbboot <- quantile(q, p)
    Q.oss <- QLR(y)
    if (Q.oss>Qsboot(y, p)) conts <- conts+1
    if (Q.oss>Qbboot) contb <- contb+1
    if (Q.oss>Qmboot(y, p)) contm <- contm+1
}

contm/s

contb/s

conts/s

```

Per applicare il bootstrap a serie storiche AR(2), simulate con parametri autoregressivi $\phi_1 = 0.5$ e $\phi_2 = 0.4$, e di numerosità campionaria $T=50$, si è utilizzato il codice R qui riportato:

```

rm(list=ls())

library(dynlm)
library(boot)

QLR <- function(y) {
  tau_0 <- round(0.15*length(y))
  tau_1 <- round(0.85*length(y))
  Q <- NULL
  mod_0 <- dynlm(y~L(y, 1:2))
  for (i in tau_0:tau_1) {
    n1 <- i
    n2 <- length(y)-i
    D <- c(rep(0,n1), rep(1,n2))
    mod_1 <- dynlm(y~L(y, 1:2)+D+D*L(y, 1:2))
    t_C <- anova(mod_0, mod_1)$F[2]
    Q[1+i-tau_0] <- t_C
  }
  max(Q)
}

contm <- 0
conts <- 0

s <- 200
phi_1 <- 0.5
phi_2 <- 0.4
T <- 50

alpha <- 1
p <- 0.95

Qmboot <- function(y, p) {
  Q <- NULL
  n <- length(y)

```

```

mod <- arima(y, order=c(2,0,0))
phil_hat <- mod$coef[1]
phi2_hat <- mod$coef[2]
media_hat <- mod$coef[3]
residui <- na.omit(mod$res)
res.c <- residui-mean(residui)
B <- 200
cost <- (1-phil_hat-phi2_hat)*media_hat
for (i in 1:B) {
  eb <- sample(res.c, n, replace=T)
  xb <- arima.sim(list(ar=c(phil_hat, phi2_hat)),
    mean=cost, innov=eb, n)

  Q[i] <- QLR(ts(xb))
}
quantile(Q, p)
}

```

```

Qsboot <- function(x, p) {
  Q <- NULL
  n <- length(x)
  y <- x-mean(x)
  fit <- ar(y)
  pstim <- fit$order
  phi.stim <- fit$ar
  residui <- na.omit(fit$res)
  res.c <- residui-mean(residui)
  B <- 200
  for (i in 1:B) {
    if (pstim == 0) {
      yb <- sample(res.c, n, replace=T)
      xb <- yb + mean(x)
    }
    if (pstim > 0) {
      yb <- NULL
    }
  }
}

```

```

        eb <- sample(res.c, n, replace=T)
        yb[1:pstim] <- mean(y)
        for (j in (pstim+1):n) {
            yb[j] <- (phi.stim)**(yb[(j-1):(j-
                pstim)]) + eb[j]
        }
        xb <- yb + mean(x)
    }
    Q[i] <- QLR(ts(xb))
}
quantile(Q, p)
}

```

```

for (i in 1:s) {
    x <- arima.sim(list(order=c(2,0,0), ar=c(phi_1,
        phi_2)), n=T)
    y <- x+alpha
    Q.oss <- QLR(y)
    if (Q.oss>Qmboot(y, p)) contm <- contm+1
    if (Q.oss>Qsboot(y, p)) conts <- conts+1
}

```

contm/s

conts/s

Bibliografia

- Andrews, D.W.K. (1993). Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica* 61 (4), 821-856.
- Berkowitz, J. e Kilian, L. (1996). Recent developments in bootstrapping time series. *Econometric Reviews* 19 (1), 1-48.
- Bickel, P.J. e Bühlmann, P. (1997). Closure of linear processes. *Journal of Theoretical Probability* 10, 445-479.
- Bose, A. (1988). Edgeworth correction by bootstrap in autoregressions. *The Annals of Statistics* 16 (4), 1709-1722.
- Brockwell, P. e Davis, R. (1991). *Time series: theory and methods*. Springer-Verlag.
- Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli* 3, 123-148.
- Bühlmann, P. (1998). Sieve bootstrap for smoothing in nonstationary time series. *The Annals of Statistics* 26 (1), 48-82.
- Bühlmann, P. (2002). Bootstrap for time series. *Statistical science* 17 (1), 52-72.
- Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The Annals of Statistics* 14 (3), 1171-1179.
- Carlstein, E., Do, K.-A., Hall, P., Hesterberg, T. e Künsch, H.R. (1998). Matched-block bootstrap for Dependent Data. *Bernoulli* 4 (3), 305-328.

- Cecchinato, N. (2008). *Bootstrap and approximation methods for long memory processes*. Tesi di dottorato in Scienze Statistiche, Università degli Studi di Padova.
- Chatterjee, S. (1986). Bootstrapping ARMA Models: Some Simulations. *Systems, Man and Cybernetic, IEEE Transactions on* 16 (2), 294-299.
- Chow, G.C. (1960). Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica* 28 (3), 591-605.
- Dahlhaus, R. e Janas, D. (1996). A frequency domain bootstrap for ratio statistics in time series analysis. *The Annals of Statistics* 24 (5), 1934-1963.
- Davidson, R. e MacKinnon, J.G. (2005). Bootstrap Methods in Econometrics. Chapter 25 in *Palgrave Handbooks of Econometrics: Vol. 1 Econometric Theory*, ed. Patterson, K.D. e Mills, T.C., Basingstoke, Palgrave Macmillan, 812-838.
- Davison, A. ed Hinkley D. (1997). *Bootstrap methods and their application*. Cambridge university press.
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics* 7, 1-26.
- Efron, B. e Tibshirani, R.J. (1986). Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science* 1, 54-75.
- Efron, B. e Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall.

- Franco, G. e Reisen, V. (2007). Bootstrap approaches and confidence intervals for stationary and non-stationary long-range dependence. *Physica A* 375 (2), 546-562.
- Franke, J. ed Härdle, W. (1992). On bootstrapping kernel spectral estimates. *The Annals of Statistics* 20 (1), 121-145.
- Freedman, D.A. (1981). Bootstrapping Regression Models. *The Annals of Statistics* 9, 1218 -1228.
- Freedman, D.A. (1984). On bootstrapping two-stages least-squares estimates in stationary linear models. *The Annals of Statistics* 12 (3), 827-842.
- Hall, P. (1985). Resampling a coverage process. *Stochastic Processes and their Applications* 19, 259-269.
- Hall, P., Horowitz, J.L. e Jing, B.Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika* 82, 561-574.
- Härdle, W., Horowitz, J.L. e Kreiss, J.P. (2001). Bootstrap Methods for Time Series. *SFB Discussion Paper*, 1-35.
- Härdle, W., Horowitz, J.L. e Kreiss, J.P. (2003). Bootstrap methods for time series. *International Statistical Review* 71, 435-459.
- Hernandez, J.A. e Mulla, D.J. (2008). Estimating Uncertainty of Economically Optimum Fertilizer Rates. *Agronomy Journal* 100 (5), 1221-1229.
- Horowitz, J.L. (2003a). The bootstrap in econometrics. *Statistical Science* 18, 211-218.

- Horowitz, J.L. (2003b). Bootstrap methods for Markov processes. *Econometrica* 71 (4), 1049-1082.
- Kapetanios, G. e Psaradakis, Z. (2006). *Sieve bootstrap for strongly dependent stationary processes*. Working Paper No. 552, Queen Mary, University of London.
- Kreiss, J.P. (1988). *Asymptotic Inference for a Class of Stochastic Processes*, Habilitationsschrift. Faculty of Mathematics, University of Hamburg, Germany.
- Kreiss, J.P. (1992). *Bootstrap procedures for $AR(\infty)$ processes*. Springer, Berlin-Heidelberg-New York.
- Kreiss, J.P. (1999). *Residual and Wild Bootstrap for Infinite Order Autoregression*. Manoscritto non pubblicato.
- Kreiss, J.P. (2000). *Nonparametric estimation and bootstrap for financial time series*, in Chan, W., Li, W. e Tong, H. (ed.), *Statistics and Finance: An Interface*, Imperial College Press, London.
- Kreiss, J.P. e Franke, J. (1989). Bootstrapping stationary autoregressive moving-average models. *Journal of Time Series Analysis* 13 (4), 297-317.
- Kreiss, J.P. e Paparoditis, E. (2003). Autoregressive-aided periodogram bootstrap for time series. *Institute of Mathematical Statistics* 31 (6), 1923-1955.
- Künsch, H. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics* 17 (3), 1217-1241.

- Lahiri, S.N. (1992). *Resampling Methods for Dependent Data*. Springer, New York.
- Lahiri, S.N. (1999a). Second order optimality of stationary bootstrap. *Statistics & Probability Letters* 11, 335-341.
- Lahiri, S.N. (1999b). Theoretical comparison of block bootstrap methods. *The Annals of Statistics* 27, 386-404.
- Lahiri, S.N. (2003). On the Impact of Bootstrap in Survey Sampling and Small-Area Estimation. *Statistical Science* 18, 199-210.
- Li, H. e Maddala G. (1996). Bootstrapping time series models. *Econometric reviews* 15 (2), 115-158.
- Liu, R.Y. e Singh, K. (1992b). *Moving blocks jackknife and bootstrap capture weak dependence*. In Lepage, R. e Billard, L. (ed.), *Exploring the Limits of the Bootstrap*, Wiley, New York , 225-248.
- MacKinnon, J.G. (2002). Bootstrap inference in econometrics. *Canadian Journal of Economics* 35, 615-645.
- MacKinnon, J.G. (2006). Bootstrap Methods in Econometrics. *Economic Record* 82, s2-s18.
- Ouyse, R. (2008). *Finite Sample Properties of the Dependent Bootstrap for Conditional Moment Models*. School of Economics UNSW, Discussion Paper, 1-33.
- Paparoditis, E. (1992). Modelling Long Term Dependence in Measurement Errors of Plutonium Concentration. *Statistical Papers* 33, 159-170.

- Paparoditis, E. (1996). Bootstrapping Autoregressive and Moving Average Parameter Estimates of Infinite Order Vector Autoregressive Processes. *Journal of Multivariate Analysis* 57, 277-296.
- Paparoditis, E. e Politis, D.N. (1999). The local bootstrap for periodogram statistics. *Journal of Time Series Analysis* 20 (2), 193-222.
- Paparoditis, E. e Politis, D.N. (2001). Tapered block bootstrap. *Biometrika* 88, 1105-1119.
- Politis, D.N. (2003). The impact of bootstrap methods on time series analysis. *Statistical science* 18 (2), 219-230.
- Politis, D.N. e Romano, J.P. (1993). Nonparametric resampling for homogeneous strong mixing random fields. *Journal of Multivariate Analysis* 47, 301-328.
- Politis, D.N. e Romano, J.P. (1994). The Stationary Bootstrap. *Journal of the American Statistical Association* 89, 1303-1313.
- Poskitt, D. (2007). Properties of the Sieve bootstrap for fractionally integrated and non-invertible processes. *Journal of Time Series Analysis* 29 (2), 224-250.
- Quandt, R.E. (1960). Tests of the Hypothesis that a Linear Regression System Obeys Two Separate Regimes. *Journal of the American Statistical Association* 55 (290), 324-330.
- Ramsey, F. (1974). Characterization of the partial autocorrelation function. *The Annals of Statistics* 2, 1296-1301.

Shumway, R.H. e Stoffer, S.S. (2000). *Time Series Analysis and its Applications*. Springer-Verlag, New York.

Stock, J.H. e Watson, M.W. (2005). *Introduzione all'econometria*. Edizione italiana a cura di Peracchi, F., Pearson.

Swanepoel, J. e van Wyk, J. (1986). The bootstrap applied to power spectral density function estimation. *Biometrika* 73 (1), 135-141.

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B. e Farmer, J. (1992). Testing for nonlinearity in time series: the method of surrogate data. *Physica D* 58, 77-94.