UNIVERSITA' DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI "M. FANNO"

DIPARTIMENTO DI AFFERENZA RELATORE: DIPARTIMENTO DI SCIENZE STATISTICHE

CORSO DI LAUREA IN ECONOMIA

PROVA FINALE

"Combinazioni previsive: focus sull'algoritmo AFTER"

RELATORE:

PROF.SSA LUISA BISAGLIA

LAUREANDA: CHUN CHEN

MATRICOLA N. 1091576

ANNO ACCADEMICO 2017 – 2018

INDICE

INTRODUZIONE

Negli ultimi anni si è visto come i dati abbiano iniziato ad assumere sempre più importanza, tanto che spesso si sente il termine “Big Data”, termine con la quale la società odierna indica appunto, l’enorme quantità di dati dalla quale persone o aziende riescono a trarre informazioni utili per le previsioni o analisi.

Proprio per questa esplosione di dati disponibili la *Statistica* è diventata un ambito che non possiamo più ignorare. Anzi andrebbe studiata con attenzione in quanto è capace di rivoluzionare e apportare vantaggi non indifferenti alla realtà ormai così dinamica.

Un ramo importante della statistica si interessa delle *serie storiche*, che definiamo come una successione di dati numerici, nella quale ogni dato è associato ad un particolare istante od intervallo nel tempo. L’analisi di queste serie è molto importante in quanto ci consente di definire l’evoluzione di un fenomeno nel tempo.

La selezione del modello per una data serie storia, può avvenire in diversi modi: criteri automatici di selezione (AIC, BIC, HQ, ecc..), verifiche di ipotesi e analisi grafiche. Tuttavia, questa selezione spesso si dimostra instabile, portando quindi ad una variabilità estremamente alta la previsione o stima finale.

In questo elaborato si presenterà il confronto tra la selezione e combinazione dei modelli, definendo i casi in cui conviene utilizzare il primo metodo e casi in cui sarebbe opportuno ricorrere al secondo.

Per la combinazione di modelli più precisamente ci si focalizzerà sull’utilizzo dell’algoritmo AFTER (*Aggregated Forecast Through Exponential Reweighting*), che combina le previsioni dai singoli modelli candidati e grazie ad un’appropriata assegnazione di pesi ad ognuna di queste previsioni, produce un *output* con accuratezza migliore rispetto alle previsioni ottenute da singoli modelli.

Nel primo capitolo verranno presentate brevemente le nozioni base, indispensabili per la comprensione di tale elaborato. Si inizierà con la definizione di serie storiche per approfondire poi i processi ARMA e ARIMA, ed infine verranno introdotti i criteri di selezione.

Nel secondo capitolo verrà spiegata l'instabilità della selezione del modello e l'introduzione della combinazione. In più verranno brevemente accennati altri metodi di combinazione, diversi dall'AFTER. Saranno introdotti anche le definizioni di ANMSEP, NMSEP e ASEP.

Nel terzo capitolo, il focus verrà spostato sulle simulazioni. In particolare, nella prima parte si dimostra che il "vero" modello, cioè il modello da cui sono stati simulati i dati, non corrisponda al modello migliore. Da sottolineare che si può parlare di "vero" modello solo in un contesto di simulazione, poiché nella realtà si può parlare solo di approssimazione del processo. Nella seconda parte invece si presentano modi per analizzare la stabilità della selezione dei modelli.

Nel quarto capitolo verrà presentato algebricamente l'algoritmo AFTER e spiegato il suo funzionamento.

Nel quinto capitolo si descrive un esperimento di simulazione per confermare quanto detto nei capitoli precedenti. In particolare, le analisi verranno svolte con: modelli semplici, modelli AR con ordini differenti, modelli ARIMA con ordini differenti.

Nel sesto capitolo ci sono rispettivamente due analisi *ex novo*, su dati reali per verificare il funzionamento dell'algoritmo in contesti applicativi. Il primo dataset riporta dati annuali del prezzo del rame e il secondo riguarda un dataset con dati annuali sulla somma delle temperature giornaliere (in gradi Celsius) al di sotto dello zero.

CAPITOLO 1

1.1. SERIE STORICHE

Uno degli obiettivi principali dell'analisi delle *serie storiche* è quello di prevedere il comportamento di un determinato fenomeno basandosi sul suo comportamento passato. Sono presenti due tipi di approcci per l'analisi delle serie storiche:

- Approccio classico: prevede sia una parte deterministica $f(\mathrm{t})$(composta da trend, ciclo e stagionalità) sia una parte stocastica u_t che tuttavia, è considerata come una componente residuale che si distribuisce come un *white noise.*
  Si definisce come:
  - Trend (o componente tendenziale): l'andamento riferito nel lungo periodo della serie, "…ed è determinato prevalentemente da fenomeni quali lo sviluppo e l'evoluzione strutturale del sistema economico, che per loro natura si manifestano con gradualità e lentamente nel tempo." (Di Fonzo, T., & Lisi, F. 2007, p.32).
    Solitamente i valori del trend sono esprimibili tramite funzione e spesso si impone anche la condizione di monotonia, per evidenziare la tendenza crescente o decrescente del fenomeno nel periodo considerato.
  - Ciclo (o componente ciclica): oscillazioni di lungo periodo, attorno al trend, scaturite da espansioni o contrazioni del sistema economico. Rispetto al trend la componente ciclica è molto meno evidente. Ci sono varie interpretazioni di tali andamenti che trovano tutte un punto comune, cioè quello delle "fluttuazioni" (l'alternanza delle fasi ascendenti e discendenti).
  - Stagionalità (o componente stagionale): "…movimenti del fenomeno nel corso dell'anno che […] tendono a ripetersi in maniera pressoché analoga nel medesimo periodo (mese o trimestre) di anni successivi." (Di Fonzo, T., & Lisi, F. 2007, p.33).
    Inoltre, un tema molto delicato è quello della stima e depurazione della stagionalità in quanto questa rappresenta un elemento di disturbo che può oscurare l'effettivo movimento ciclico della serie.
  - Errore (o componente stocastica): Elemento che contiene tutte le variabili di entità trascurabile che non vengono considerate in $\mathrm{Y_t}$.

La componente stocastica, quindi è l'elemento che porta irregolarità (positive o negative) nelle serie. In aggiunta, in questo approccio si ipotizza che siano prodotte da un processo aleatorio, implicando quindi la propensione di queste irregolarità ad annullarsi tra di loro, compensandosi l'un l'altro. Per queste le definiamo *white noise,* cioè un processo con media zero e varianza σ_ε^2.

Questo approccio quindi fa assunzioni molto forti, poiché afferma che al netto delle componenti deterministiche la componente stocastica sia un processo *white noise* eliminando quindi l'ipotesi di correlazioni seriali dei residui.

- Approccio moderno: che sposta il focus sulla componente stocastica u_t , la quale viene ora vista come un processo a componenti correlate (quindi dipendente da Y_t e u_t passati).
  Anche l'obiettivo cambia, qui si punta infatti ad individuare il modello probabilistico che meglio descrive l'evoluzione del fenomeno più che la stima delle componenti della serie.

Un'altra nozione importante per la discussione dell'elaborato è la definizione di processo stocastico:

- Dati uno spazio di probabilità $(\Omega, \mathcal{F}, \mathrm{P})$ e "…uno spazio parametrico T, si definisce come processo stocastico una funzione finita a valori reali di $t \epsilon T$ e $\omega \epsilon \Omega$ tale che, per ogni t, $Y_t(\omega)$ è una funzione misurabile di ω." (Di Fonzo, T., & Lisi, F. 2007, p.160)

Per la modellazione della componente stocastica inoltre è essenziale che tale processo sia stazionario, cioè:

- Media e varianza entrambe costanti, quindi non dipendenti dal tempo
- Le autocovarianze dipendono dalla distanza temporale tra due variabili e non dagli istanti $\mathrm{t}_1, \mathrm{t}_2$

Esistono due definizioni di stazionarietà, in senso forte e in senso debole. Quella più usata è la seconda. Diremo che un processo Y_t è stazionario in senso debole se:

- $\mathrm{E}[\mathrm{Y}] = \mu \qquad \forall \mathrm{t},$
- $\mathrm{Cov}[\mathrm{Y_t}, \mathrm{Y_{t+k}}] = \gamma_k \qquad \forall \mathrm{t}, \quad \forall \mathrm{k}.$

Questa è anche la definizione più usata, poiché oltre che essere meno restrittiva della stazionarietà in senso forte, per molti processi è sufficiente e soddisfacente la stazionarietà di secondo ordine

1.2. PROCESSI ARMA E ARIMA

- Processo autoregressivo a media mobile – ARMA(p,q)
  È una combinazione dei modelli AR(p) e MA(q):
  - Modello AR(p) viene definito come processo autoregressivo e si scrive come:

$$\emptyset(\mathrm{B})\mathrm{Y_t} = \emptyset_0 + \varepsilon_\mathrm{t}$$

    - B è l'operatore ritardo che trasforma la serie $\mathrm{y_t}$ nella serie ritardata di un periodo $\mathrm{y_{t-1}} = \mathrm{By_t}$
    - $\emptyset(\mathrm{B}) = \left(1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_\mathrm{p} B^p\right)$
    - $\{\varepsilon_t\}$ rappresenta un processo *white noise* con media zero e varianza σ_ε^2
    - $\emptyset_\mathrm{i}(\mathrm{i} = 0, \dots, \mathrm{p})$ parametri costanti

    Quindi la variabile Y_t è pensata come il risultato di una somma pesata sia di valori passati che di uno shock casuale contemporaneo.

  - Il modello MA(q) invece è un processo a media mobile e si scrive come:

$$Y_t = \theta(B)\varepsilon_t \,.$$

    - $\theta(\mathrm{B}) = 1 - \theta_1 \mathrm{b} - \cdots - \theta_\mathrm{P} \mathrm{b}^\mathrm{P}$
    - $\theta_j(j = 0, \dots, q)$ parametri costanti

    Dove anche qui $\{\varepsilon_t\}$ rappresenta un processo *white noise* con media zero e varianza σ_ε^2.

  Mentre nei processi AR e MA si possono avere un numero molto elevato di parametri, con il modello misto ARMA possiamo avere una rappresentazione più parsimoniosa.
  Partendo da un processo white noise $\{\varepsilon_t\}$, si dice che $\{\mathrm{Y_t}\}$ è un processo autoregressivo a media mobile di ordine (p,q) se si può scrivere come:

$$\mathrm{Y_t} - \sum_{i=1}^{p} \emptyset_i Y_{t-i} = \emptyset_0 + \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}.$$

- se $q = 0$ allora si ha un modello AR(p)
- se $p = 0$ si ha un modello MA(q).

Questo processo è stato proposto da Box e Jenkins (1976) come classe standard per il trattamento delle serie storiche.

Può essere espresso con l'operatore di ritardo:

$$\emptyset(\mathrm{B})\mathrm{Y_t} = \emptyset_0 + \theta(\mathrm{B})\varepsilon_\mathrm{t}$$

Si dice stazionario quando tutte le p radici dell'equazione caratteristica $\emptyset(B) = 0$ sono in modulo maggiori di uno.

- Processo integrato a media mobile – ARIMA (p,d,q)

  Questo processo è un'estensione del processo ARMA, il quale però considera processi non stazionari omogenei di grado d che possono essere trasformati in stazionari grazie alle opportune trasformazioni.

  Sia $\{\varepsilon_t\}$ un processo white noise con media zero e varianza σ_ε^2.

  Si indichi con X_t la d-esima differenza di Y_t, $X_t = (1-B)^d Y_t$. Si dice che Y_t è un processo autoregressivo integrato a media mobile di ordine (p,d,q) se $\{X_t\}$ è un processo ARMA(p,q)

$$\mathrm{X_t} = (1-\mathrm{B})^\mathrm{d}\mathrm{Y_t},$$

$$X_t = \emptyset_0 + \sum_{i=1}^{p} \emptyset_i X_{t-i} + \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}.$$

  Scrittura con l'operatore di ritardo è:

$$\emptyset(\mathrm{B})(1-\mathrm{B})^\mathrm{d}\mathrm{Y_t} = \emptyset_\mathrm{o} + \theta(\mathrm{B})\varepsilon_\mathrm{t}.$$

  Se si ha:

  - $d = q = 0$ allora la classe di processi ARIMA (p,d,q) diventa un processo AR(p)
  - $\mathrm{p} = \mathrm{d} = 0$ allora si ottiene la classe dei processi di MA(q)
  - $\mathrm{d} = 0$ $\mathrm{Y_t}$ è un processo ARMA
  - $\mathrm{p} = \mathrm{q} = 0$ è un processo *random walk*
  - $\mathrm{p} = \mathrm{q} = \mathrm{d} = 0$ allora si ha un processo *white noise*

Per la stima dei parametri di un modello ARIMA uno dei metodi più usati è quello della massima verosimiglianza, il quale è estremamente efficiente sotto assunzione di normalità delle osservazioni.

Tuttavia, questo metodo presenta un problema rilevante, infatti è possibile aumentare la verosimiglianza di un modello attraverso l'aggiunta di parametri, ma così facendo si può incorrere nell' *overfitting.*

L'*overfitting* o eccesso di ottimizzazione è la situazione in cui un modello statistico si adatta ai dati osservati abusando di un numero eccessivo di parametri. Per evitare questo scenario si introducono diversi criteri di selezione che possiamo ricavare dalla scrittura generica:

$$-log(massima\ verosimiglianza) + penalità$$

1.3. CRITERI DI SELEZIONE DEL MODELLO

I criteri di selezione del modello, aiutano a limitare gli effetti collaterali dell'*overfitting*, ovvero la pratica di inserire troppi parametri nel modello per fare in modo che questo si adatti bene ai dati. Il fatto che il modello si adatti bene ai dati non implica anche che faccia una buona previsione, infatti più si inseriscono parametri superflui più aumenta la variabilità della stima.

I criteri di selezione presi in considerazione in questo elaborato sono:

- AIC (*Akaike Information Criterion*): con penalità $p + q$
- BIC ((*Bayesian Information Criterion*): con penalità $[(p + q) \cdot ln(n)]/2$
- HQ (*Hannan-Quinn Information Criterion*): con penalità $(p + q) \cdot ln\big(ln(n)\big)$

I criteri qui sopra presentati verrano utilizzati per gli studi empirici di questo elaborato.
Le proprietà di questi criteri in sintesi sono:

- BIC e HQ sono consistenti nel senso che la probabilità di selezionare il vero modello tende ad 1 (se il vero modello è nella lista dei candidati)
- AIC invece è distorto, e tende a sovraparametrizzare, infatti è il criterio che attribuisce la penalità minore tra i tre.

  Per questo motivo l'AIC tenderà sempre a preferire un modello di ordine maggiore a quello vero.

CAPITOLO 2

2.1 INSTABILITÀ NELLA SELEZIONE DEL MODELLO

Per l'analisi e la previsione di serie storiche, l'uso dei modelli statistici, come per esempio il modello ARIMA, si è dimostrato molto efficace nelle applicazioni relative alle procedure di previsione.

Tuttavia, una complicazione che si può riscontrare è quello della scelta del "miglior" modello.

Esistono sono vari approcci per affrontare il problema soprammenzionato, e in questo elaborato si andrà ad esaminare il metodo che si serve dei criteri AIC, BIC, HQ e AICc.

Uno dei maggiori svantaggi della selezione del modello è l'*instabilità*.

Infatti, se si hanno poche osservazioni o comunque un numero moderato, i modelli simili sono difficili da distinguere, e anche i valori dei criteri di selezione dei modelli non daranno risultati attendibili.

In questo caso è inappropriato scegliere il modello con il valore dei *criteria* più basso, poiché un lieve cambiamento dei dati potrebbe portare alla scelta di un modello differente.

Quindi quando le previsioni basate sulla selezione dei modelli hanno una variabilità estremamente alta, Yang (2001c) ha mostrato come teoreticamente una combinazione di modelli sia la soluzione migliore.

In questa tesi verrà analizzato un algoritmo (AFTER) che combina le previsioni dai singoli modelli candidati e che grazie ad un'appropriata assegnazione di pesi ad ognuno di queste previsioni, fornirà un *output* con accuratezza migliore.

Questo metodo è stato studiato in maniera teorica da Yang (2001b) e mostra come la procedura di combinazione raggiunga il miglior tasso di convergenza offerta dai candidati modelli individuali.

Tuttavia, non è chiaro se questi risultati teorici possano descrivere quello che succederebbe in una reale applicazione.

Per questa analisi il focus è incentrato sui modelli ARIMA, simulazioni e dati reali verranno usati per mettere a confronto la combinazione dei modelli e la selezione in termini di accuratezza nella previsione.

2.2 METODI DI COMBINAZIONI DI MODELLI GIÀ PROPOSTI.

Bates e Granger (1969) affermano che in una combinazione lineare di due previsioni, con l'assegnazione dei pesi in modo ottimo (minimizzando quindi la varianza), in generale si avrà un errore quadratico medio più piccolo delle due singole previsioni individuali.

Potrebbe sembrare strano come la combinazione con una previsione peggiore possa portare ad un miglioramento, tuttavia con la giusta assegnazione dei pesi è possibile.

Inoltre, l'unico caso in cui non si vedrà nessun miglioramento è quando una previsione è già ottima, quindi il peso dato a questa sarà 1 mentre alle altre verrà dato 0.

Vent'anni dopo Granger (1989) evidenzia come nel suo primo lavoro le due previsioni fossero basate implicitamente sullo stesso set di informazioni.

Continua poi affermando che per ottenere una previsione migliore tramite combinazione, non solo le previsioni individuali devono basarsi su diversi *dataset* ma è anche necessario che ognuno di questi sia ottimale per il loro set di informazioni.

Anche Clement e Hendry (1998) hanno dichiarato che quando le previsioni sono basate su modelli (econometrici) che hanno accesso allo stesso set di informazioni, allora la combinazione di queste previsioni può non essere ottima.

Proprio per questo è meglio individuare singoli modelli per ricavare un modello "preferito" che contenga le caratteristiche utili dei modelli di origine.

In questo elaborato tuttavia attribuiamo importanza anche alla combinazione di previsioni che provengono da modelli simili in quanto così facendo si riduce la variabilità.

Inoltre, mentre le precedenti combinazioni di modelli tentano di migliorare le previsioni individuali, l'AFTER punta al miglioramento della performance del miglior modello candidato.

Il termine "incertezza del modello" è stato usato per esprimere il problema di identificazione del modello corretto, tuttavia nel nostro caso l'identificazione riguarda il modello "migliore" (valutato secondo la funzione di perdita: errore quadratico medio di previsione).

Si parte dall'ipotesi che il "vero" modello potrebbe esserci come non esserci nella lista dei modelli candidati e che anche se ci fosse il "vero" modello potrebbe non corrispondere al modello "migliore".

Sebbene questo elaborato si focalizzi sulla combinazione dei modelli, tuttavia riconosce l'importanza della selezione dei modelli in quanto identificando il "vero" modello, questo ci permette di capire le relazioni tra le variabili.

In una regressione lineare si nota che è consigliata la selezione dei modelli quando un modello è preferito in maniera evidente e quindi c'è poca instabilità nella selezione.

Nelle serie storiche quando la selezione dei modelli è stabile, la combinazione non apporta nessun tipo di miglioramento.

2.3 ALCUNE OSSERVAZIONI

Valutazioni di accuratezza della previsione.

Si assume che per $i \geq 1$, la distribuzione condizionata di Y_i date le osservazioni precedenti abbia media m_i e varianza v_i^2.

Con $\widehat{Y_i}$ definiamo un valore predetto di Y_i basato sull'informazione passata Y^{i-1}

Allora l'errore quadratico medio per la previsione con orizzonte temporale pari a uno è uguale a:

$$\mathrm{E}\left(\mathrm{Y_i} - \hat{Y}_\mathrm{i}\right)^2$$

Tale quantità essere utilizzata come misura di performance per la previsione di Y_i .

Questo può essere scomposto in:

$$E_i\left(Y_i - \hat{Y}_i\right)^2 = \left(m_i - \hat{Y}_i\right)^2 + v_i^2$$

Per v_i^2 intendiamo la varianza condizionata, che non si può controllare ed è presente sempre senza distinzioni del modello utilizzato. E poiché la varianza condizionata è la stessa per tutte le previsioni, allora la si può togliere per misurare la performance delle procedure di previsione.

Quindi definiamo la perdita netta come:

$$\mathrm{L}\left(\mathrm{m_i}, \hat{Y}_i\right) = \left(m_i - \hat{Y}_i\right)^2$$

E come rischio netto:

$$\mathrm{E}\left(\mathrm{m_i} - \hat{Y}_\mathrm{i}\right)^2$$

consideriamo le medie degli errori quadratici medi:

$$ANMSEP(\delta, n_0, n) = \frac{1}{n - n_0 + 1} \sum_{i=n_0+1}^{n+1} E\left(\mathrm{m_i} - \hat{Y}_\mathrm{i}\right)^2$$

Con δ si indica il procedimento di previsione che fornisce le previsioni di $\hat{Y}_1, \hat{Y}_2$, ...,

Per il caso particolare di $n_0 = n$ ANMSEP diventa NMSEP(δ, n) che è semplicemente il valore netto dell'errore quadratico medio.

In questo elaborato ci focalizzeremo sul ANMSEP per la comparazione della selezione dei modelli con la combinazione dei modelli nelle simulazioni.

Per l'analisi di dati reali si prende come riferimento:

$$ASEP(\delta, n_0, n) = \frac{1}{n - n_0} \sum_{i=n_0+1}^{n} \left(Y_i - \widehat{Y}_\iota\right)^2$$

Con n_0 indichiamo una frazione non troppo piccola di n. E al contrario della quantità teorica ANMSEP, la quantità ASEP può essere calcolata.

CAPITOLO 3

3.1 CONFRONTO TRA DUE MODELLI ANNIDATI

Se si considerano due modelli annidati, è quasi automatico implementare come approccio un test di verifica d' ipotesi per vedere quale dei due è quello che può aver generato i dati.

Esempio 0: Vengono considerati due modelli:

Modello 1: $Y_n = e_n$;
Modello 2: $Y_n = \alpha Y_{n-1} + e_n$

Con $\{e_n\}$ i.i.d normalmente distribuita con media zero e varianza σ^2. Si assume che $0 < \alpha < 1$. Come si vede dalla formula il modello 1 è annidato nel modello 2 quando $\alpha = 0$.

Con il modello 1 il valore quadratico medio della previsione dell'errore è minimizzato quando $\hat{Y}_{n+1} = 0$.

Se α non è uguale a zero allora il NMSEP viene scritto come:

$$E\left(\hat{Y}_{n+1} - \alpha Y_n\right)^2 = \alpha^2 E(Y_n^2)$$

Per il modello 2, indichiamo come $\hat{\alpha}_n$ la stima di α e la previsione è $\tilde{Y}_{n+1} = \hat{\alpha}_n Y_n$, quindi scriviamo come NMSEP:

$$E\left(\tilde{Y}_{n+1} - \alpha Y_n\right)^2 = E(\hat{\alpha}_n - \alpha)^2 Y_n^2$$

Per comparare i due modelli possiamo utilizzare la seguente funzione:

$$\frac{\alpha^2 E Y_n^2}{E(\hat{\alpha}_n - \alpha)^2 Y_n^2}$$

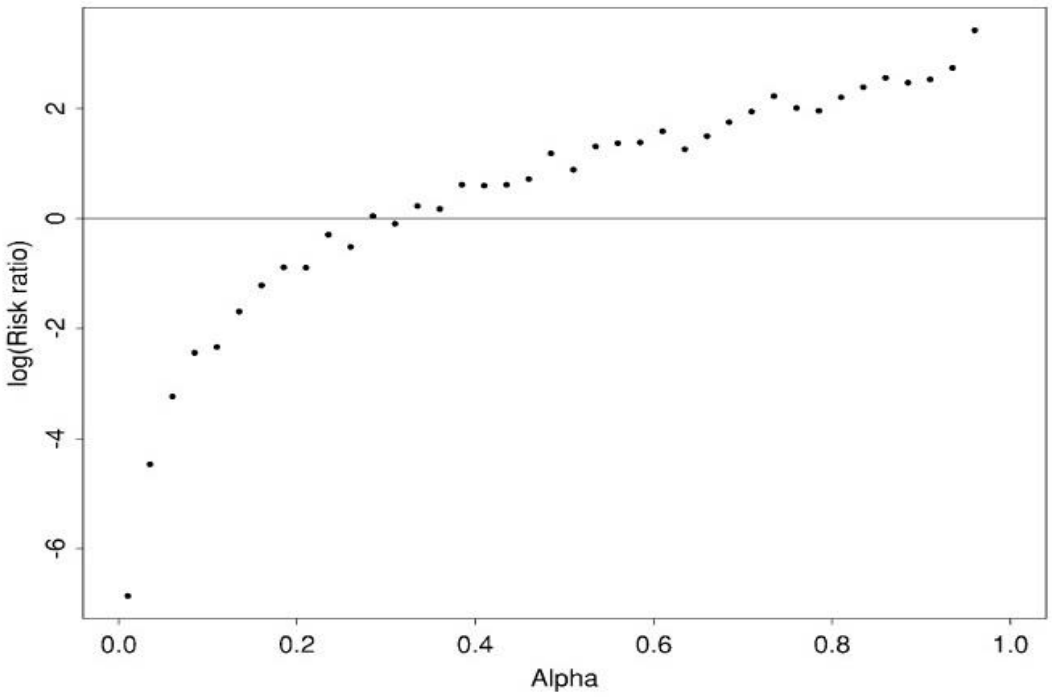


Fig.1. Confronto dei modelli giusti e sbagliati nella predizione

Nella figura 1, con $n = 20$ e $\sigma^2 = 1$, si ha il grafico che mette a confronto il rapporto dei due modelli (su una scala logaritmica), con Alpha.

Si sono svolte grazie alle simulazioni Monte Carlo 1000 repliche per ogni scelta di α.

Dal grafico si nota che quando α è piccola (meno di circa 0.3), il modello sbagliato ha una *performance* migliore, tuttavia con l'aumentare di α è preferibile il vero modello.

Non è tuttavia chiaro quale sia la relazione tra il livello di significatività del test e quale dei due modelli sia migliore nella predizione.

3.2 MISURARE LA STABILITA' DEI METODI DI SELEZIONE DEI MODELLI.

Quando ci sono incertezze nel trovare il modello migliore, metodi alternativi per ottenere previsioni, quali la combinazione, vengono presi in considerazione. Persiste tuttavia il problema di come misurare la stabilità della selezione del modello o del risultato predetto.

Qui si propongono due tipi di approcci:

- STABILITÀ SEQUENZIALE:

Prendendo un metodo di selezione del modello, un modo di misurare la stabilità della selezione è quella di esaminare come cambia l'ordine scelto dal criterio di selezione quando si utilizzano campioni di grandezza differente.

Supponiamo che in base a tutte le osservazioni $\{Y_i\}_{i=1}^{n}$ sia stato selezionato il modello $\hat{k}_n$. Poniamo come K la percentuale delle volte in cui lo stesso modello viene selezionato:

$$\mathrm{K} = \frac{\sum_{j=n-L}^{n-1} I_{\widehat{\{k}_j = \hat{k}_n\}}}{L}$$

La formula ci permette di intuire che se il criterio di selezione è stabile, la rimozione di alcune osservazioni non dovrebbe portare a troppi cambiamenti.

L è un numero intero, ed è il numero di osservazioni in testa alla serie (osservazioni più recenti) che dedichiamo al calcolo della stabilità sequenziale. Deve essere scelto con cura. Da una parte si vuole scegliere un L piccolo così che i problemi di selezione per ogni j in $\{\mathrm{n}-\mathrm{L},\dots,\mathrm{n}-1\}$ siano simili al problema originale. Tuttavia, bisogna ricordarsi che per ottenere un K abbastanza affidabile L non deve essere troppo piccolo.

Se un metodo è instabile sequenzialmente, il criterio ha difficoltà nel selezionare un modello ben preciso, allora si può affermare che è anche improbabile che scelga il modello migliore.

In aggiunta, bisogna notare come K non risolva il problema della selezione, in quanto un modello può essere anche stabile ma avere performance scarse.

- STABILITÀ RISPETTO AD UNA PERTURBAZIONE:

Un altro approccio per misurare la stabilità di un modello di selezione è attraverso la perturbazione.

L'idea alla base è: se un processo statistico è stabile allora una piccola perturbazione dei dati non dovrebbe modificare di troppo il risultato.

Considerando un criterio di selezione dei modelli per comparare i modelli ARMA che hanno come formula:

$$\emptyset(\mathrm{B})\mathrm{Y_i} = \theta(\mathrm{B})\mathrm{e_i}$$

Con:

- $\emptyset(B) = 1 - \emptyset_1 B - \cdots - \emptyset_p B^p$
- $\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$

Identifichiamo $\hat{p}$ e $\hat{q}$ gli ordini selezionati dal criterio, quindi:

- $\widehat{\emptyset}(B) = 1 - \widehat{\emptyset}_1 B - \cdots - \widehat{\emptyset}_{\hat{p}} B^{\hat{p}}$
- $\hat{\theta}(B) = 1 + \hat{\theta}_1 B + \cdots + \hat{\theta}_{\hat{q}} B^{\hat{q}}$

Dove $\widehat{\theta_\iota}$ e $\widehat{\emptyset}_i$ sono parametri stimati basandoci sui dati. Generiamo quindi una serie storica seguendo il modello:

$$\widehat{\emptyset}(B)W_i = \hat{\theta}(B)\eta_\mathrm{i}$$

Dove η_i sono i.i.d. $\sim N(0, \tau^2\hat{\sigma}^2)$ con $\tau > 0$ e $\hat{\sigma}^2$ una stima della varianza basata sul modello selezionato.

Si consideri ora $\tilde{Y}_i = Y_i + W_i$ per ogni $1 \leq i \leq n$ e si applichino i criteri per la selezione dei modelli ai nuovi dati $\{\tilde{Y}_i\}_{i=1}^n$.

Se il criterio di selezione del modello è stabile per i dati quando τ è piccola, il nuovo modello selezionato risulta essere più o meno come il primo e la previsione corrispondente non dovrebbe cambiare troppo.

- Stabilità nella selezione

  Per ogni τ segniamo la percentuale delle volte in cui si sceglie nuovamente il modello scelto precedentemente. Infine, facciamo una tabella mettendo in confronto la percentuale (delle volte) con τ.

  Se la percentuale diminuisce rapidamente in τ, questo indica che la procedura di selezione è instabile anche con una piccola perturbazione.

- Instabilità nelle previsioni

  La stabilità nella selezione non cattura necessariamente la stabilità nelle previsioni poiché modelli differenti possono avere la stessa bontà di performance nelle previsioni.

Quindi per ogni τ inseriamo la deviazione media della nuova previsione da quella originaria in base alla varianza stimata (usando il modello selezionato inizialmente), basandoci su 100 repliche.

Quindi noi facciamo la media di

$$\frac{|\tilde{Y}_{n+1} - \hat{Y}_{n+1}|}{\hat{\sigma}}$$

Su 100 perturbazioni indipendenti di misura τ dove $\tilde{Y}_{n+1}$ è ottenuto con un'applicazione della procedura di selezione sui dati perturbati, e $\hat{Y}_{n+1}$ e $\hat{\sigma}$ sono stime basate sui dati.

Un modo ragionevole per calcolare l'instabilità della previsione è quanto velocemente cresce questa quantità (la media riportata sopra) all'aumentare di τ.

- ESEMPI CON DATA SET

Si prendano in considerazione:

- *dataset* 1 (Wei, G. C., & Tanner, M. A. 1990, p.446) e
- *dataset* 3 (Makridakis, S., Wheelwright, S., & Hyndman, F. 1998, Capitolo 1).

  - DATA SET 1

I criteri di selezione prendono in considerazione i modelli AR dall'ordine 1 a 5. Wei suggerisce Ar (1) come modello per il dataset.

La fig. 2 mostra le perturbazioni e la stabilità (di selezione e previsione)

C'è una piccola incertezza per quanto riguarda la stabilità della selezione quando aumenta la perturbazione, infatti nei grafici (a destra) si nota come AIC, BIC, e HQ si allontanino da 1.0 con l'aumentare di τ.

Quindi in questo caso non c'è nessun vantaggio nell'usare l'algoritmo AFTER, questo è dimostrato anche dal confronto degli ASEP(n_0, n) che sono rispettivamente:

- 0.249 per AIC, BIC, e HQ
- 0.254 per l'algoritmo AFTER

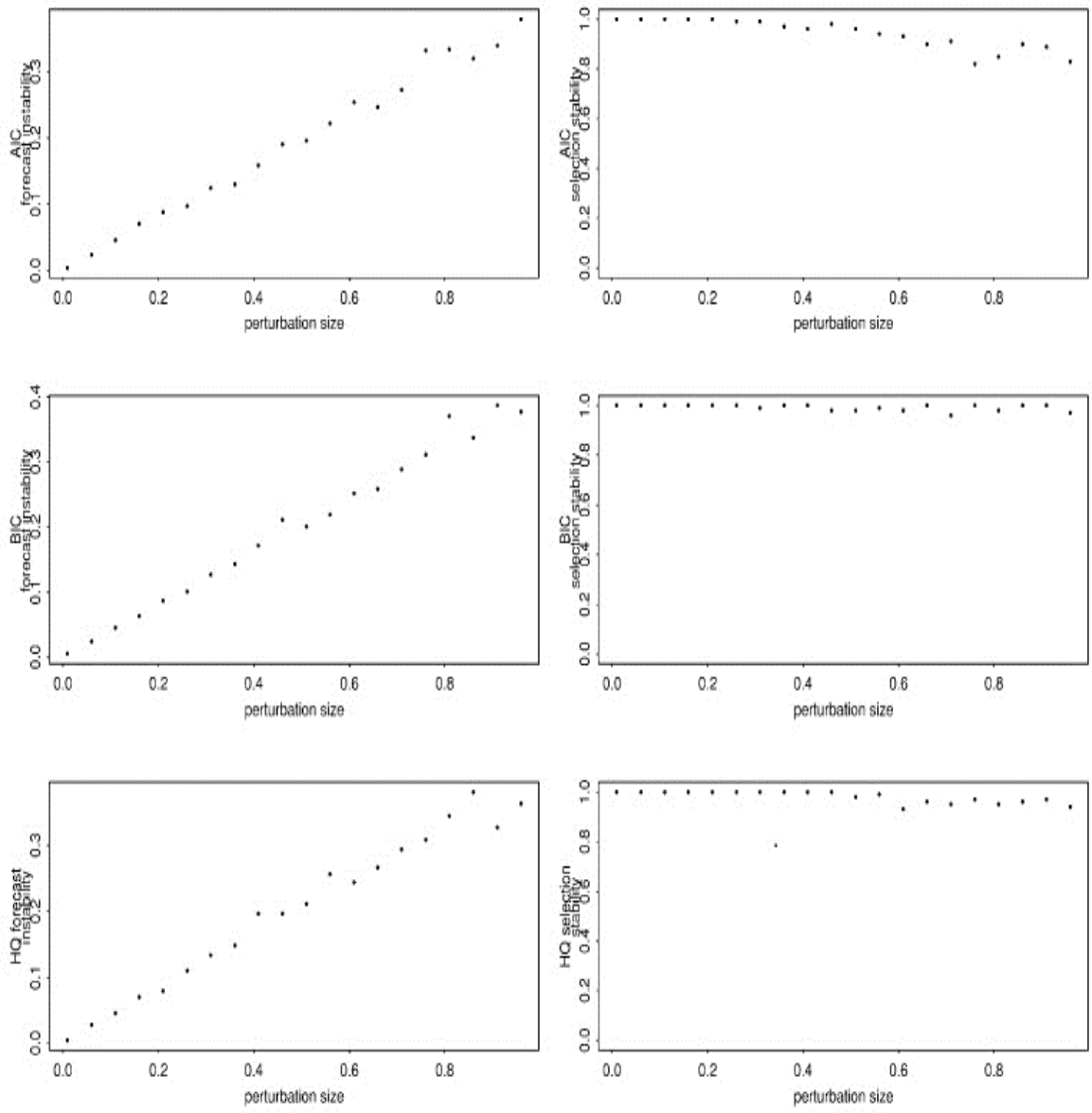


Fig.2 Stabilità della previsione e selezione per il data set 1.

Da sottolineare come sebbene la stabilità della selezione rimanga nei dintorni di 1.0 (soprattutto se si guardano i criteri BIC e HQ), è necessario notare come comunque un valore alto di K sia ininfluente sulla stabilità della previsione, la quale diminuisce con l'aumentare delle perturbazioni.

- DATA SET 3 (È USATA UNA TRASFORMAZIONE LOGARITMICA)

La figura 3 mostra come siano instabili i metodi della selezione dei

modelli.

In tutti e tre i criteri l'instabilità nella previsione incrementa molto velocemente con τ. La tabella di perturbazione può essere molto utile per decidere se combinare o selezionare in una situazione pratica.

Infatti, paragonato ai risultati del data set 1, l'instabilità delle previsioni del data set 3 è molto alta con il metodo della selezione dei modelli, soprattutto per quanto riguarda i criteri BIC e HQ.

In questo caso invece si nota come l'algoritmo AFTER sia preferibile alla selezione, realtà evidenziata anche dagli ASEP(n_0, n) che sono rispettivamente:

- 704.2 (AIC), 813.4 (BIC), 785.2 (HQ)
- 635.3 (AFTER)

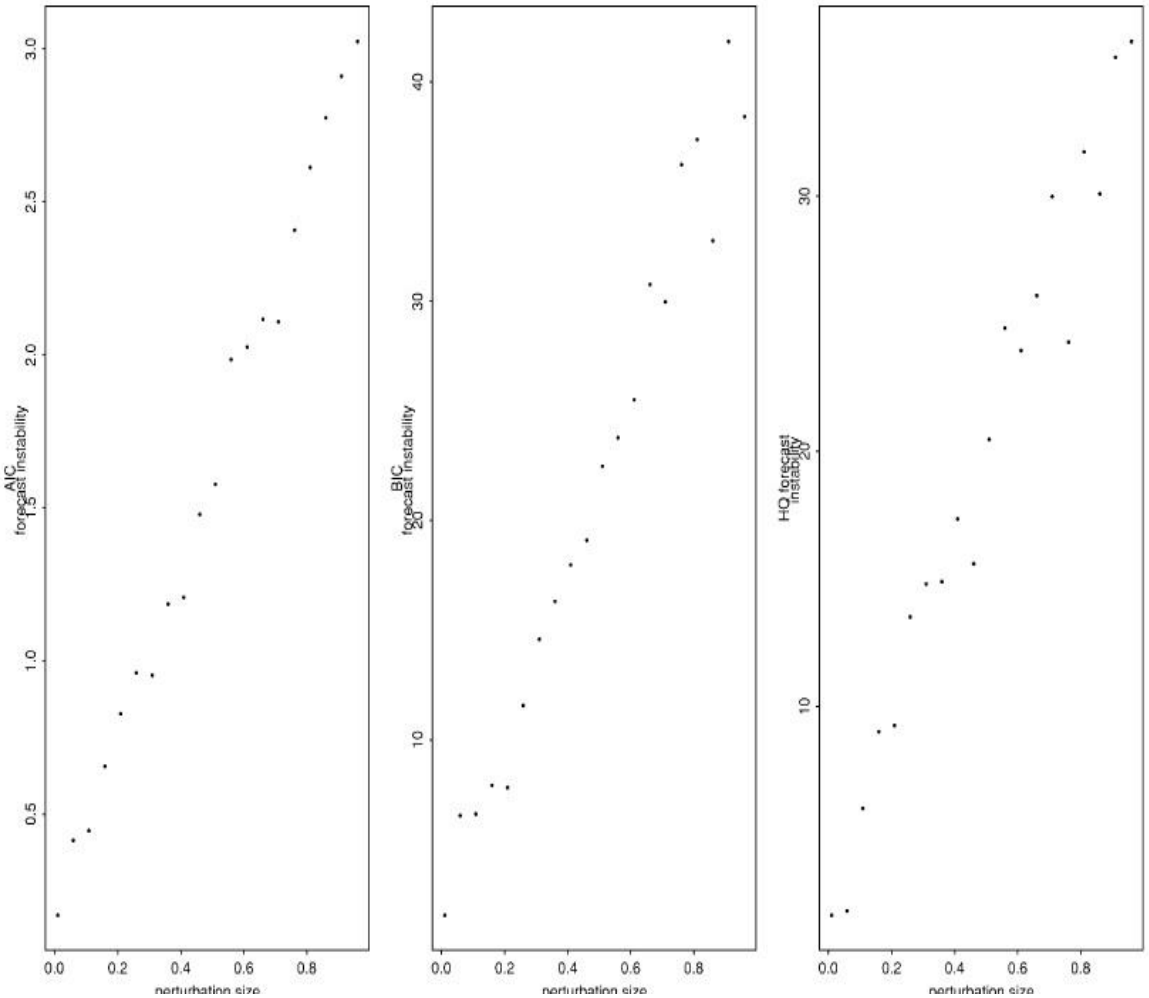


Fig.3 instabilità nella previsione per il data set 3.

CAPITOLO 4

4.1 ALGORITMO AFTER

Si assume che la distribuzione condizionata di Y_i, dato $Y^{i-1} = y_{i-1}$, sia gaussiana per $i \geq 1$ con una media condizionata pari a m_i e una varianza condizionata pari a v_i.

Si assume che per ogni procedura di previsione δ_j, assieme alla previsione $\hat{Y}_{j,n}$, si ottiene da $\hat{Y}^{n-1}$ anche $\hat{v}_{j,n}$ la quale è una stima di v_n.

Yang(2001b) ha proposto un algoritmo chiamato AFTER per la combinazione di diverse previsioni e ne ha studiato le di convergenza.

In questo elaborato applichiamo il modello AFTER(*Aggregated Forecast Through Exponential Reweighting*) nel caso in cui molteplici modelli dello stesso tipo siano usati per la previsione.

Per combinare le procedure di previsione $\Delta = \{\delta_1, \delta_2, \dots, \delta_j\}$ per ogni tempo n, l'AFTER guarda a tutte le loro performance passate e assegna loro i seguenti pesi:

$W_{j,1} = 1/J$ e per $n \geq 2$ si ha:

$$W_{j,n} = \frac{W_{j,n-1}\hat{v}_{j,n-1}^{-1/2}\exp\{-[(y_{n-1} - \hat{y}_{j,n-1})^2/2\hat{v}_{j,n-1}]\}}{\sum_{j'=1}^{J} W_{j',n-1}\hat{v}_{j',n-1}^{-1/2}\exp\{-[(y_{n-1} - \hat{y}_{j',n-1})^2/2\hat{v}_{j',n-1}]\}}.$$

Le previsioni poi vanno combinate con la seguente formula:

$$\hat{y}_n^* = \sum_{J=1}^{J} W_{j,n}\hat{y}_{j,n}$$

Si nota che $\sum_{J=1}^{J} W_{j,n} = 1$ per $n \geq 1$, cioè la sommatoria dei pesi deve sempre essere 1 se abbiamo $n \geq 1$, inoltre dipendono solo dalle previsioni e realizzazioni passate di Y.

Inoltre, va evidenziato come per ogni osservazione aggiuntiva i pesi nelle previsioni candidate vengano aggiornati.

CAPITOLO 5

5.1 SIMULAZIONI

- DUE SEMPLICI MODELLI

In questa sezione studiamo la differenza tra selezione e combinazione di modelli.

Quello che segue è il risultato da una simulazione eseguita da Yang(2004) con $n = 20$ con 1000 replicazioni dei due modelli annidati presentati nella sezione (3.1)

Per l'applicazione del metodo AFTER, impostiamo $n_0 = 12$ e analizziamo in base alle prime 12 osservazioni.

	AIC	BIC	HQ	AICc
$\alpha = 0.01$	0.157	0.094	0.135	0.144
$\alpha = 0.50$	0.580	0.463	0.557	0.568
$\alpha = 0.91$	0.931	0.884	0.921	0.924

Tabella 1. Percentuale di selezionare il vero modello

	True	Wrong	AIC	BIC	HQ	AICc	AFTER
$\alpha = 0.01$							
$n_0 = 12$	0.116	0.000	0.036	0.025	0.036	0.033	0.029
	(0.03)	(0.000)	(0.003)	(0.002)	(0.003)	(0.002)	(0.001)
$n_0 = 20$	0.094	0.000	0.028	0.017	0.025	0.028	0.027
	(0.005)	(0.000)	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)
$\alpha = 0.50$							
$n_0 = 12$	0.158	0.339	0.211	0.234	0.211	0.216	0.158
	(0.005)	(0.007)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)
$n_0 = 20$	0.124	0.332	0.168	0.192	0.173	0.170	0.156
	(0.008)	(0.015)	(0.011)	(0.012)	(0.011)	(0.011)	(0.010)
$\alpha = 0.91$							
$n_0 = 12$	0.387	5.057	0.944	1.140	0.934	0.988	0.552
	(0.011)	(0.195)	(0.053)	(0.068)	(0.053)	(0.003)	(0.022)
$n_0 = 20$	0.320	5.019	0.614	0.721	0.615	0.615	0.596
	(0.020)	(0.230)	(0.045)	(0.071)	(0.060)	(0.049)	(0.048)

Tabella.2 Selezione di modello e combinazione dei modelli a confronto con ANMSEP e NMSEP

Nella tabella 1 sono riportate le percentuali di selezionare il vero modello con i criteri AIC, BIC, HQ e AICc ai rispettivi α valori.

La tabella 2 invece presenta il ANMSEP del vero modello, del modello sbagliato, AIC, BIC, HQ, AICc (una variante dell'AIC contenente una correzione per piccoli campioni) e AFTER.

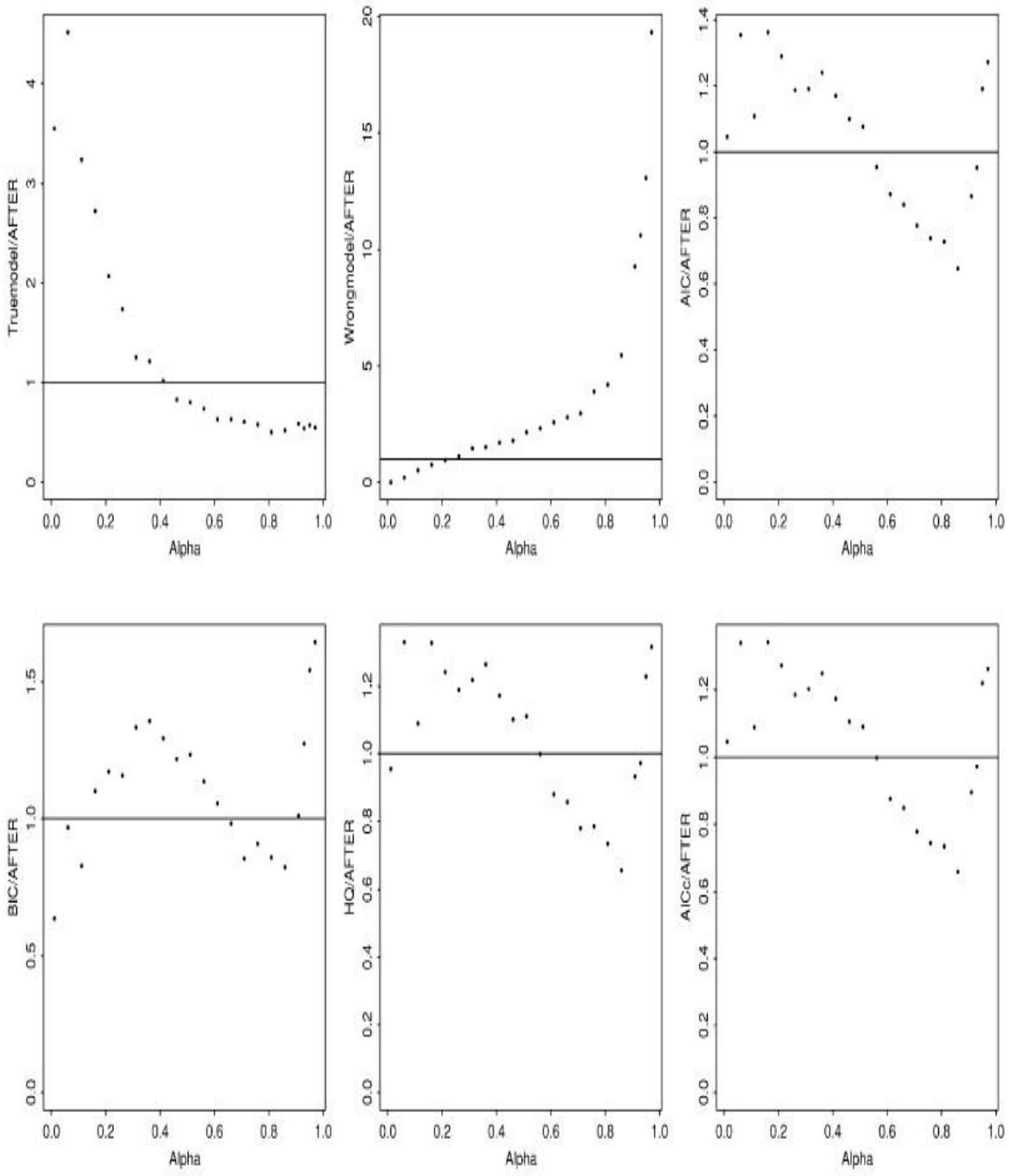


Fig.4. Confronto della selezione del modello con AFTER per i due casi di modelli in NMSEP

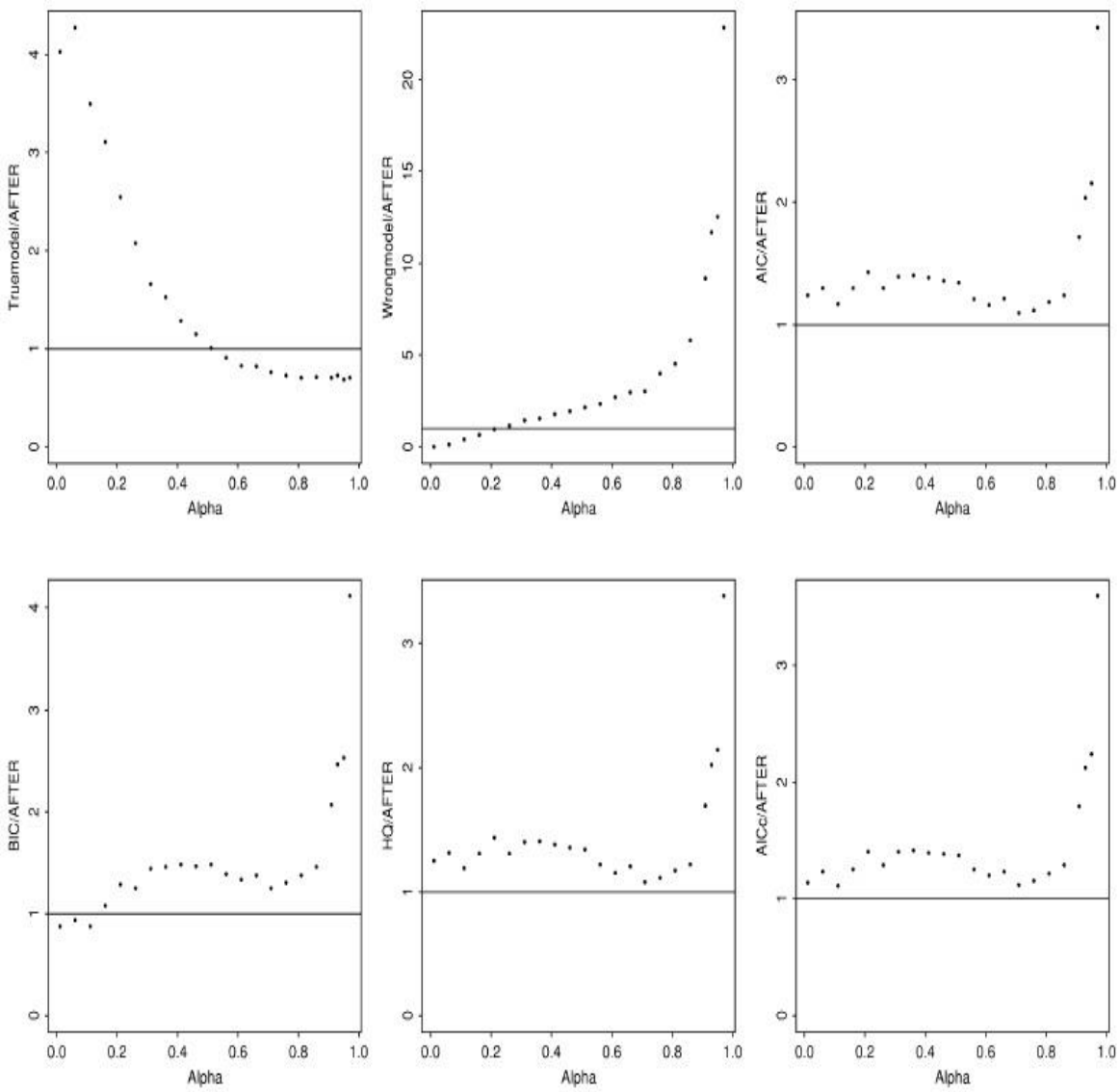


Fig.5. Confronto della selezione dei modelli con AFTER per i due casi di modello in ANMSEP

Le figure 4 e 5, invece, confrontano i criteri di selezione dei modelli con l'AFTER in termini di NMSEP e ANMSEP, entrambi basati su 1000 simulazioni.

Nella fig.4 (NMSEP) quando α non è troppo piccolo, ma comunque più piccolo di 0.55, l'AFTER ha una performance migliore dei criteri di selezione del modello.

Ci si aspetterebbe che quando si ha α è piccolo o grande, i criteri di selezione del modello scelgano con un'alta probabilità il modello migliore (che non è necessariamente il vero modello).

Tuttavia, sorprendentemente l'AFTER è in vantaggio per un α vicino ad 1.

La possibile spiegazione per questo caso è che, visto che il modello sbagliato ha una *performance* così scarsa, anche se si ha una probabilità davvero bassa nel selezionarla con la selezione dei modelli tuttavia questo danneggia lo stesso il rischio medio (ANMSEP e NMSEP).

In contrasto, l'algoritmo AFTER assegnando i pesi adeguati riesce a ridurre l'influenza del modello sbagliato.

E' interessante notare come quando si prenda in considerazione l'ANMSEP, l'AFTER sembra più avvantaggiato, forse perché essendoci poche osservazioni, si ha più instabilità nella selezione.

Per confermare quanto appena detto, si è svolto una simulazione nella stabilità della selezione con due perturbazioni: $\tau = 0.2$ e $\tau = 0.6$.

Si è aggiunto alla serie un ulteriore campionamento con $n = 50$ per evidenziare il trend più chiaramente.

Per $\alpha = 0.01$ la stabilità della selezione non cambia troppo.

Le medie della stabilità della selezione per $\alpha = 0.5\ e\ 0.91$ sono presentati nella seguente tabella.

Si hanno 100 replicazioni con 1000 perturbazioni di due misure diverse ognuna.

	AIC	BIC	HQ	AICc
$\alpha = 0.5$				
$n = 12$	0.867	0.876	0.863	0.871
	0.778	0.824	0.771	0.794
$n = 20$	0.872	0.856	0.865	0.871
	0.802	0.832	0.805	0.809
$n = 50$	0.959	0.920	0.948	0.962
	0.928	0.883	0.912	0.926
$\alpha = 0.91$				
$n = 12$	0.880	0.874	0.874	0.878
	0.775	0.782	0.780	0.790
$n = 20$	0.957	0.926	0.955	0.954
	0.921	0.888	0.918	0.912
$n = 50$	1.000	1.000	1.000	1.000
	1.000	0.999	1.000	1.000

Tabella 3. Stabilità della selezione con due perturbazioni per 3 campionamenti.

Da questa tabella possiamo vedere come la stabilità della selezione aumenti con l'aumentare delle osservazioni, soprattutto se guardiamo $\alpha = 0.91$, e questa stabilità diminuisce l'influenza della perturbazione.

	True	Wrong
$\alpha = 0.01$	0.374	0.626
$\alpha = 0.50$	0.519	0.481
$\alpha = 0.91$	0.793	0.207

Tabella 4. Media dei pesi nei modelli con AFTER.

Questa tabella riporta la media dei pesi che AFTER da ai due modelli e si vede come con l'aumentare di α aumenti anche il peso dato al modello giusto.

- MODELLI AR CON ORDINI DIFFERENTI

I modelli candidati sono AR con ordini fino all'8. Consideriamo il rischio ANMSEP con $n = 50$ e due n_0 (20 e 50) per la procedura di confronto delle previsioni.
Cinque casi sono stati scelti come esempi per sottolineare il confronto tra selezione e combinazione.

- Caso 1. Il vero modello è AR(1) con coefficiente 0.8.
- Caso 2. Il vero modello è AR(2) con coefficienti (0.278, 0.366)
- Caso 3. Il vero modello è AR(3) con coefficienti (0.7, -0.3, 0.2)
- Caso 4. Il vero modello è AR(4) con coefficienti (0.386, -0.36, 0.276, -0.227)
- Caso 5. Il vero modello è AR(7) con coefficienti (-0.957, 0.0917, 0.443, 0.834, 0.323, -0.117, 0.0412)

La varianza dell'errore è 1.

Confrontiamo le *performance* di AIC, BIC e HQ con AFTER.

Sono considerati due versioni di AFTER:

1. Il primo considera il peso uniforme a priori per prevedere Y_{21} per poi passare ai pesi dati dall'algoritmo AFTER. (AFTER)
2. Essendoci inizialmente 20 osservazioni, è possibile considerare anche l'opzione di iniziare ad assegnare pesi basandoci sui primi 15 e poi aggiornarli con le osservazioni rimaste.
   Così facendo l'uso del peso uniforme a priori avrà meno influenza sulle previsioni. (AFTER2)

La simulazione è basata su 1000 replicazioni e viene riportata nella tabella sottostante.

	AIC	BIC	HQ	AFTER	AFTER2	SA
Case 1						
$n_0 = 20$	0.164	0.139	0.151	0.161	0.145	0.210
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
$n_0 = 50$	0.107	0.084	0.089	0.092	0.084	0.131
	(0.007)	(0.006)	(0.006)	(0.006)	(0.005)	(0.007)
Case 2						
$n_0 = 20$	0.162	0.174	0.163	0.139	0.139	0.149
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
$n_0 = 50$	0.103	0.122	0.105	0.097	0.097	0.104
	(0.005)	(0.007)	(0.006)	(0.005)	(0.005)	(0.005)
Case 3						
$n_0 = 20$	0.167	0.163	0.164	0.137	0.137	0.149
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
$n_0 = 50$	0.121	0.131	0.125	0.097	0.102	0.098
	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)
Case 4						
$n_0 = 20$	0.194	0.223	0.205	0.152	0.166	0.137
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)
$n_0 = 50$	0.141	0.184	0.157	0.128	0.141	0.102
	(0.007)	(0.008)	(0.008)	(0.006)	(0.006)	(0.005)
Case 5						
$n_0 = 20$	0.813	0.865	0.835	0.745	0.767	0.833
	(0.012)	(0.013)	(0.013)	(0.012)	(0.012)	(0.014)
$n_0 = 50$	0.604	0.771	0.681	0.593	0.658	0.606
	(0.039)	(0.042)	(0.039)	(0.034)	(0.036)	(0.034)

Tabella 5. Confronto tra metodo di selezione e combinazione con i modelli AR

Osserviamo che, per il caso 1, il criterio di selezione del modello ha difficoltà a trovare il modello migliore e che neanche l'AFTER si rivela vantaggioso.

Infatti, si nota che ha una *performance* decisamente peggiore del BIC per $n_0 = 15$ e $n_o = 50$.

Per tutti gli altri casi l'AFTER ha una performance migliore dei criteri di selezione.

Per quanto riguarda il confronto tra i due AFTER, tranne nel caso 1, AFTER2 ha una performance peggiore dell'AFTER.

- MODELLI *RANDOM*

Per dimostrare che i vantaggi di AFTER non sono atipici, il confronto ora verrà fatto in una circostanza *random*.

Il vero processo generatore dei dati è stato scelto casualmente tra i modelli AR di ordine compreso tra 1 e 8. I coefficienti associati al modello così estratto, sono stati generati secondo una distribuzione uniforme [-10,10].

Si ottengono così 110 processi.

	AIC	BIC	HQ
Median loss ratio	1.347	1.286	1.247
Risk ratio	1.663	1.678	1.578
	(0.092)	(0.133)	(0.099)

Tabella 6. Confronto tra i due metodi con i modelli AR: caso random

Il confronto di questa tabella si basa sull'esaminare il tasso del rischio netto per $n = 100$ e $n_0 = 50$ con 100 repliche per ogni processo.

L'assegnazione dei pesi per il modello AFTER inizia al tempo 50.

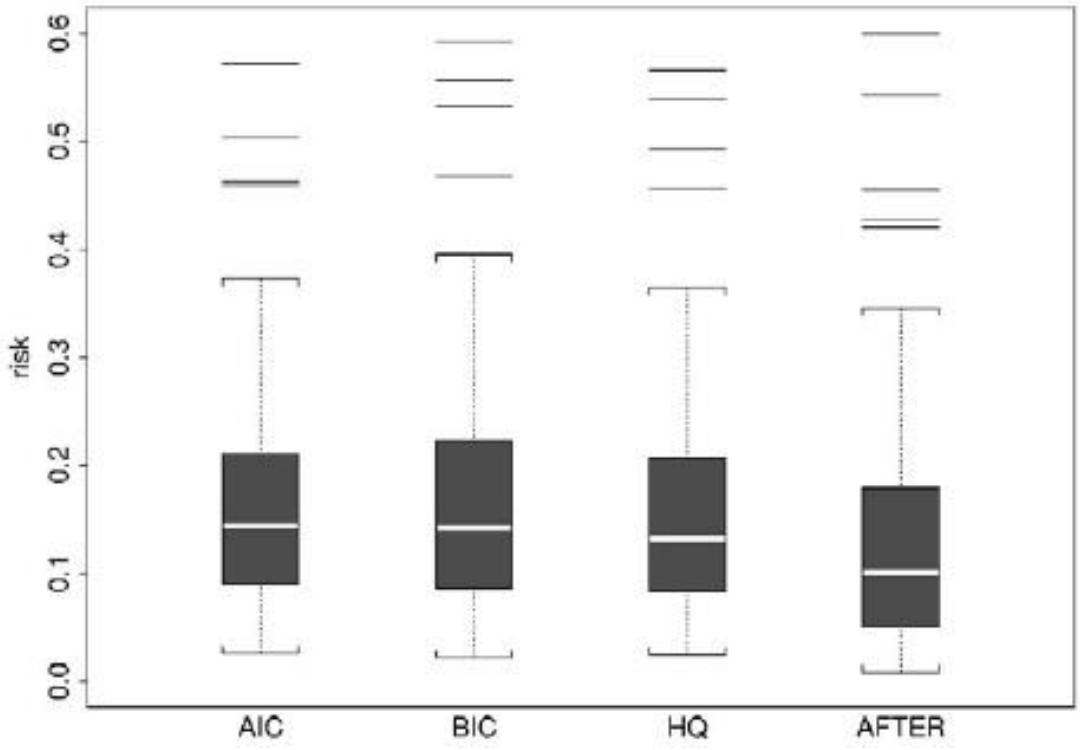


Fig.6 Rischi di AIC, BIC, HQ e AFTER con modelli AR random

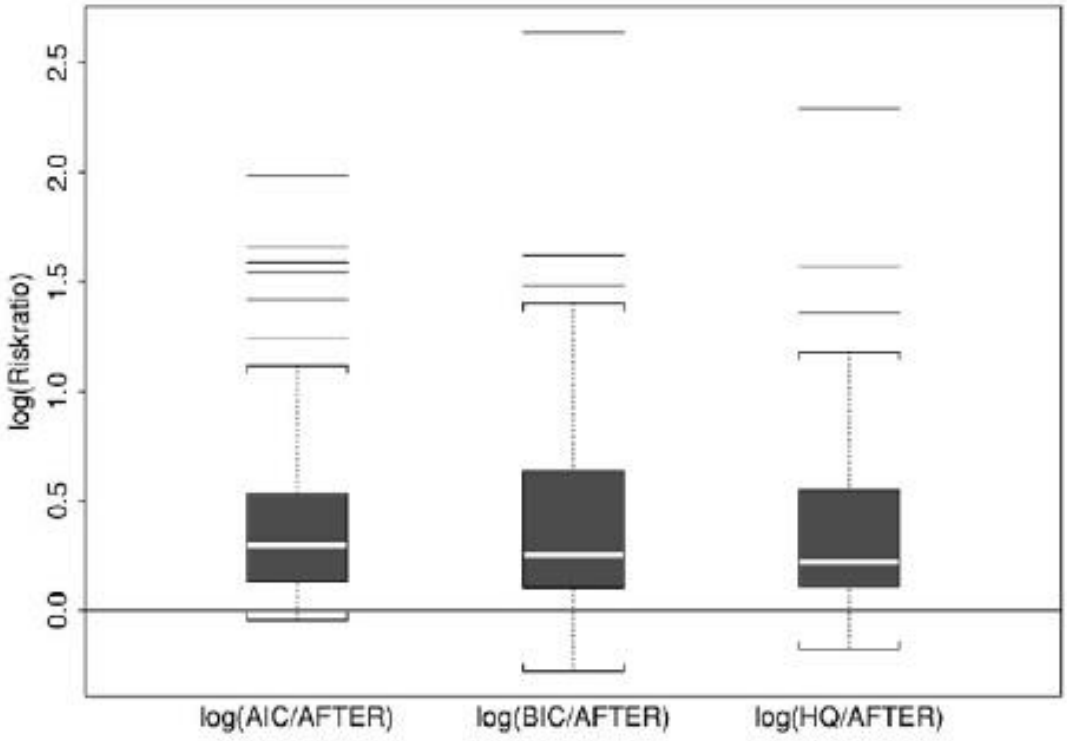


Fig.7. Tasso di rischio di AIC, BIC, HQ confrontati con AFTER. Modelli AR random

Le tabelle mostrano chiaramente come l'AFTER abbia performance migliori per quanto riguarda la previsione.

- MODELLI ARIMA CON ORDINI DIFFERENTI.

Il confronto qui segue la stessa metodologia vista sopra, ma viene fatto con i modelli ARIMA con $n = 100$ e $n_0 = 50$ con 100 replicazioni.

Il processo generatore è un ARIMA (2,1,2), coefficienti AR sono (0.8, 0.1) e i coefficienti MA sono (0.6, 0.1).

I modelli candidati sono ARIMA(p,d,q) con $p, q = 1, \dots, 8$ e $d = 1,0$

La tabella riportata sotto mostra come AFTER abbia una performance migliore in media.

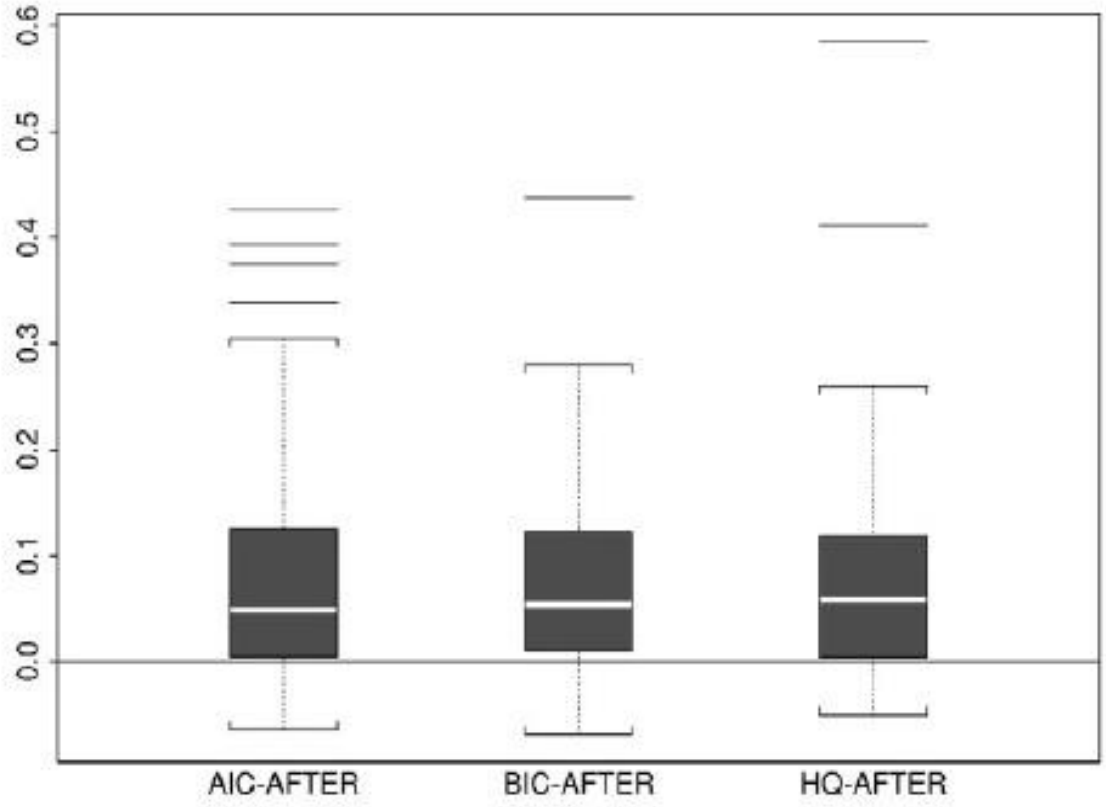


Fig.8. Confronto AIC, BIC, HQ a paragone con AFTER. Modelli ARIMA.

CAPITOLO 6

In questo capitolo saranno eseguiti, con il linguaggio di programmazione R, analisi *ex novo* di due dataset:

- Prezzo del Rame (Fonte: *Makridakis, Wheelwright and Hyndman (1998)*)
- Temperature Negative in Inverno (Fonte: *Hipel and McLeod (1994)*

In particolare, il focus di questo capitolo è incentrato su:

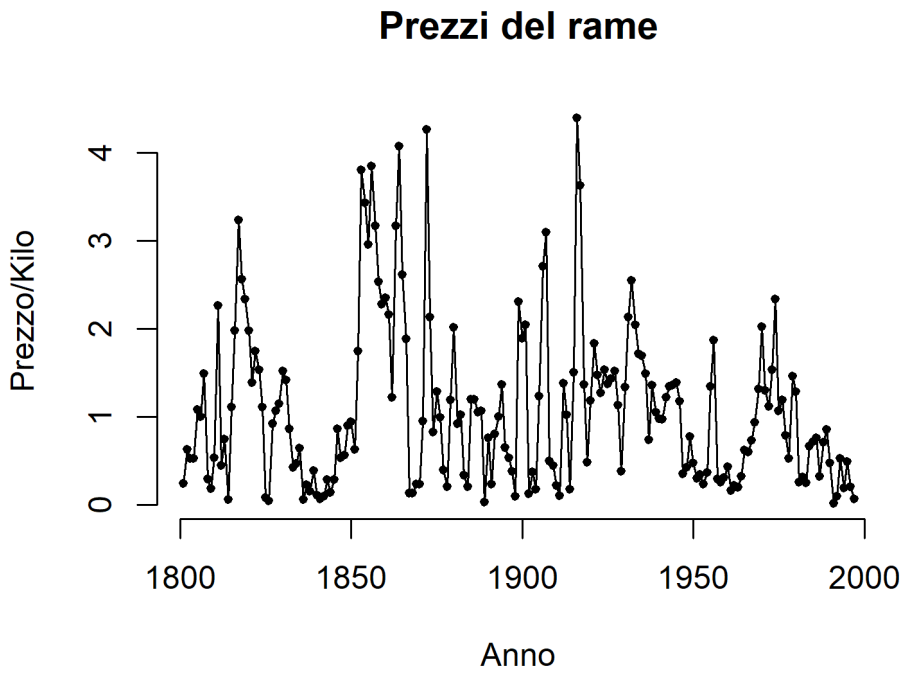
- Analisi dell'instabilità sequenziale
- *Performance* dei criteri di selezione confrontati con il risultato che invece ci fornisce l'AFTER attraverso il calcolo dell'ASEP.

6.1 ANALISI SU DATI REALI

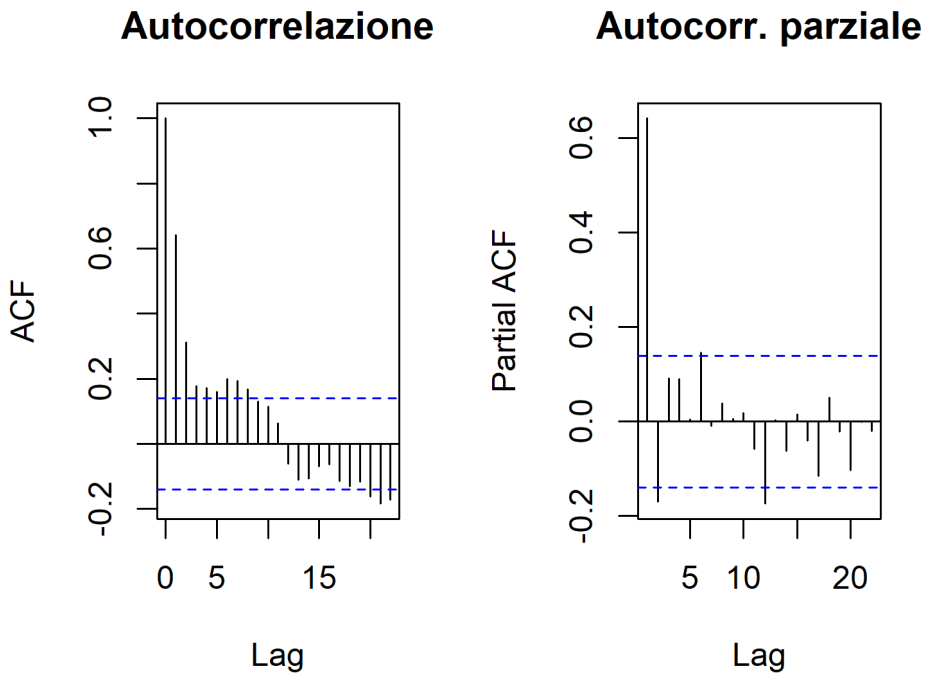
- DATASET: PREZZO DEL RAME

  Il dataset considerato riporta dati annuali del prezzo del rame negli Stati Uniti, in dollari al kilo (197 osservazioni), nel periodo 1801-1997. Per eliminare l'effetto dell'inflazione il prezzo è considerato sempre in riferimento al dollaro del 1997.

```
dati <- read.csv("copper.csv", sep = ",", he = T)[,2]; dati <- ts
(dati)
t <- 1801:1997
plot(t, dati, main="Prezzi del rame", xlab="Anno",
     ylab="Prezzo/Kilo", type="l", frame.plot = F)
points(t, dati, pch=20, cex=0.8)
```



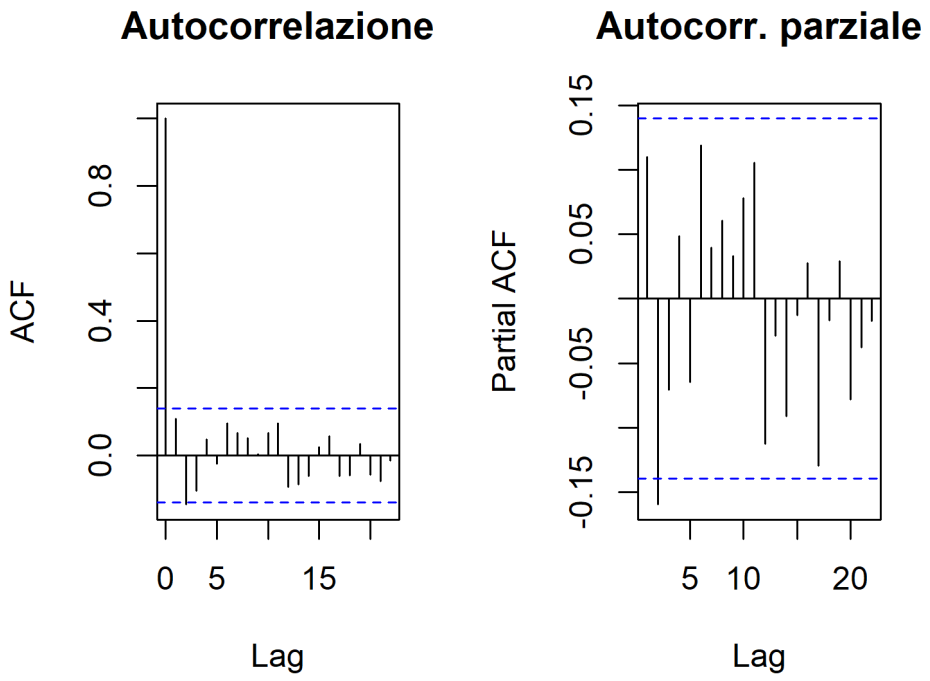

```
par(mfrow=c(1,2)); acf(dati, main=""); title("Autocorrelazione")
pacf(dati, main=""); title("Autocorr. parziale")
```



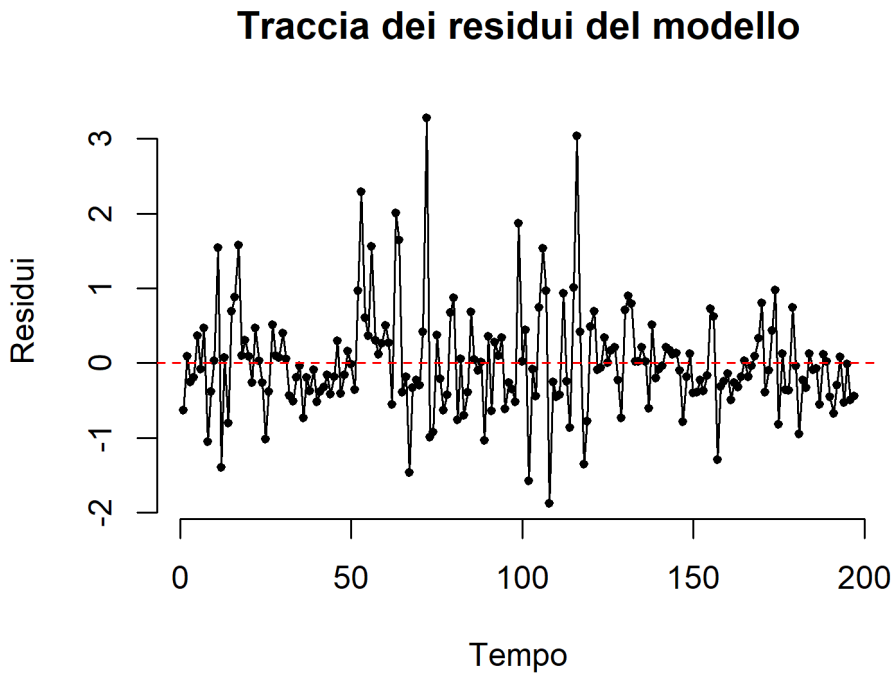
La funzione di autocorrelazione globale decresce esponenzialmente, funzione di autocorrelazione parziale presenta un ritardo significativamente diverso da 0; il processo sembra poter essere modellato con un AR(1):

```r
m <- arima(dati, c(1,0,0))

par(mfrow=c(1,2))
acf(residuals(m), main=""); title("Autocorrelazione")
pacf(residuals(m), main=""); title("Autocorr. parziale")
```




```r
par(mfrow=c(1,1))
plot(residuals(m), main="Traccia dei residui del modello",
     xlab = "Tempo", ylab = "Residui", frame.plot=F)
points(residuals(m), pch=20, cex=0.8)
abline(h=0, lty=2, col=2)
```



L'adattamento sembra soddisfacente, ma è probabile che i criteri di selezione automatica possano propendere per modelli più complessi. I candidati sono i modelli $ARIMA(p,d,q)$, $p = 0 \dots 4$, $q = 0 \dots 3$, $d = 0{,}1$:

```
data.frame(choosing.best(dati, 4,1,3))
```

	p	d	q
AIC	3	0	2
BIC	1	0	1
HQ	3	0	2

La scelta ricade invece su modelli più complessi, come evidenziato dalla tabella. L'instabilità sequenziale è calcolata con $L = 60$ e vale:

```
data.frame(matrix(round(instab.assess(dati, 60, c(4,1,3), perc =
F),3),1,3,
       dimnames=list(NULL, c("AIC","BIC","HQ"))))
```

	AIC	BIC	HQ
k	0.167	0.25	0.183

Sembra esserci molta instabilità in tutti i criteri di selezione. L'AFTER dovrebbe portare ad un miglioramento delle previsioni in questo contesto.

L'$ASEP(137{,}197)$ calcolato utilizzando i diversi algoritmi è riportato sotto:

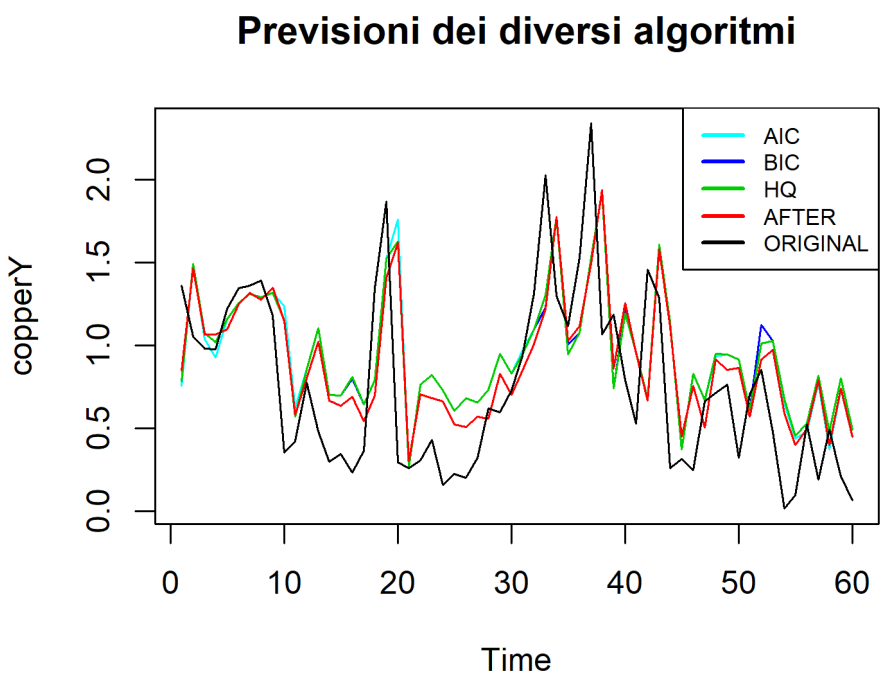
```
foo <- ICmodels(dati, 60, order=c(4,1,3))
foo2 <- AFTER(dati, 60, c(4,1,3))

copperY <- cbind(foo$yhat, foo2$yhat, foo$truth)
copperY <- ts(winterY, names=c("AIC","BIC","HQ","AFTER","ORIGINAL
"))

x <- c(foo$asep[-4], foo2$asep)
copper <- data.frame(matrix(x, 1, 4, byrow=T,
                  dimnames = list(c("ASEP"), c("AIC","BIC","HQ","
AFTER"))))
copper
```

	AIC	BIC	HQ	AFTER
ASEP	0.2190986	0.2155728	0.2160026	0.1969214

```
plot(copperY, plot.type = "single", col=5:1,
     main="Previsioni dei diversi algoritmi")
legend("topright", legend=colnames(copperY), lwd=2, col=5:1, cex=
0.7)
```

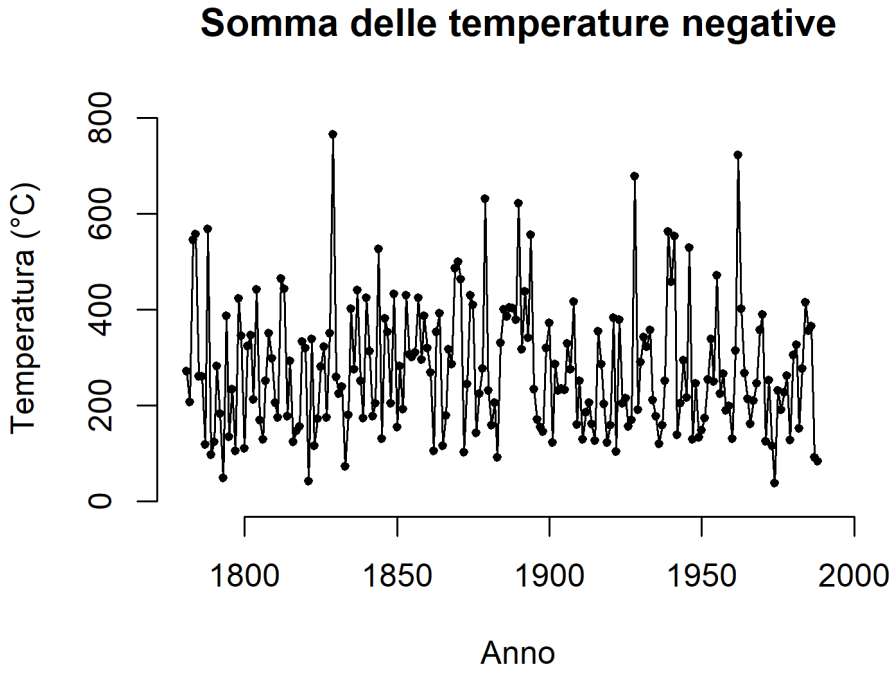


I risultati mostrano un effettivo miglioramento dell'algoritmo AFTER nelle previsioni.

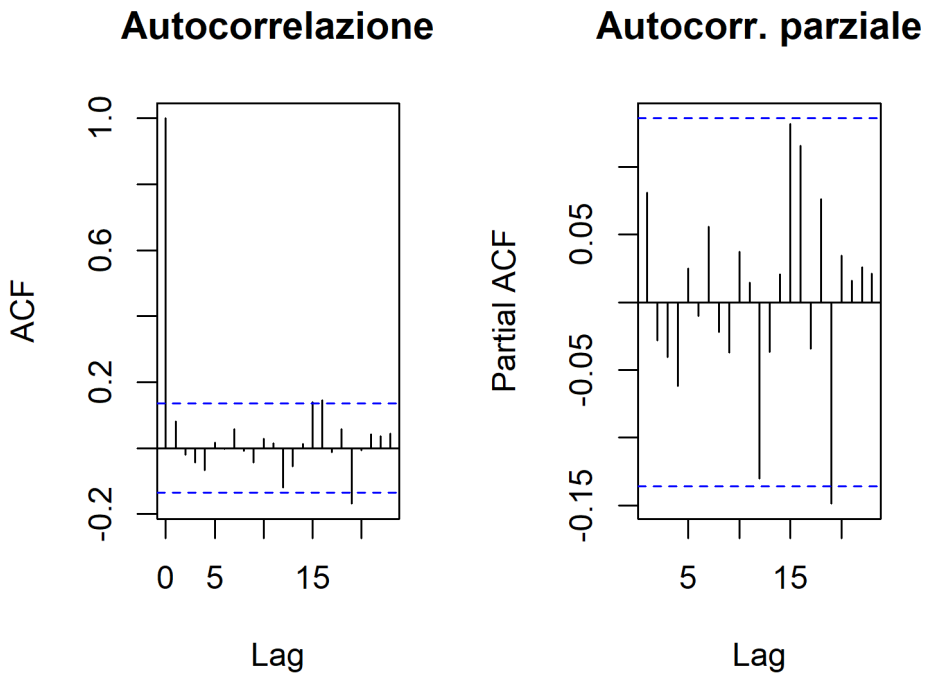
- DATASET: TEMPERATURE NEGATIVE IN INVERNO

  Il dataset considerato riporta dati annuali sulla somma delle temperature giornaliere (in gradi Celsius) al di sotto dello zero, relative al Canada, nel periodo 1781-1988. Sommando le temperature che vanno sotto zero nell'intero anno solare e ipotizzando che questo accade prevalentemente in inverno, possiamo usare questa analisi per determinare quanto gli inverno di ogni anno siano stati rigidi (208 osservazioni).

```
dati <- read.csv("winter.csv", sep = ",", he = T)[,2]; dati <- ts
(dati)
t <- 1781:1988
plot(t, dati, main="Somma delle temperature negative", xlab="Anno
",
     ylab="Temperatura (°C)", type="l", frame.plot = F,
     xlim=c(1780, 2000), ylim=c(0,805))
points(t, dati, pch=20, cex=0.8)
```

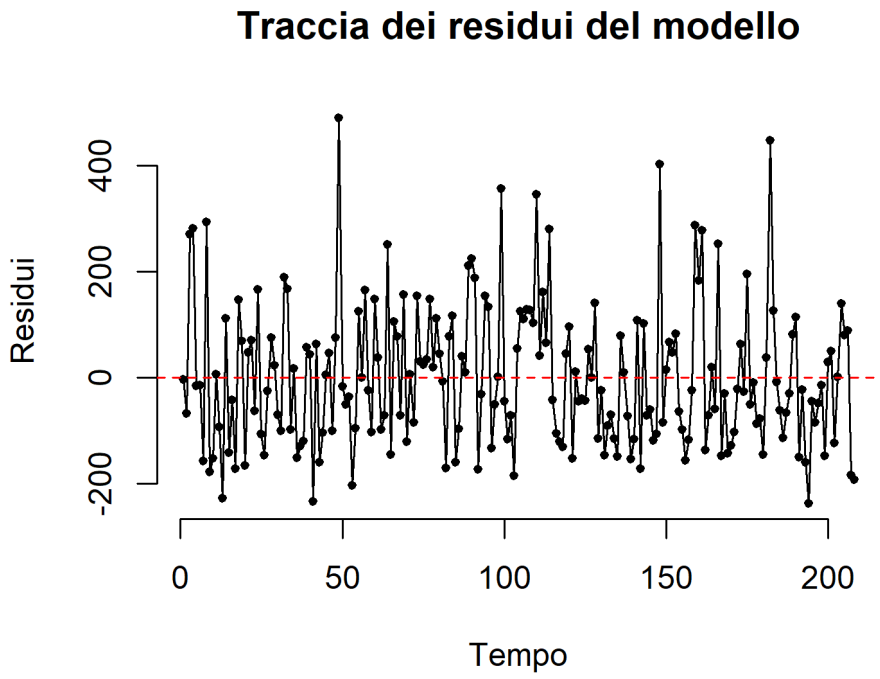



```
par(mfrow=c(1,2)); acf(dati, main=""); title("Autocorrelazione")
pacf(dati, main=""); title("Autocorr. parziale")
```




```r
m <- arima(dati, c(0,0,0))

par(mfrow=c(1,1))
plot(residuals(m), main="Traccia dei residui del modello",
     xlab = "Tempo", ylab = "Residui", frame.plot=F)
points(residuals(m), pch=20, cex=0.8)
abline(h=0, lty=2, col=2)
```



Questo caso di studio è un po' particolare, stiamo considerando un processo che si

comporta quasi come un *white noise*. A prima vista sembra quasi che non siano necessari modelli per prevedere, se non la banale media aritmetica delle osservazioni. Ma alcuni modelli più complessi potrebbero comunque presentare un adattamento migliore, almeno secondo i criteri di selezione automatica. I candidati sono i modelli $ARIMA(p,d,q)$, $p,q = 0,1,2,3$, $d = 0,1,2$:

```
data.frame(choosing.best(dati, 3,2,3))
```

	p	d	q
AIC	0	1	1
BIC	0	1	1
HQ	0	1	1

Tutti e tre i criteri selezionano un modello ARIMA(0,1,1) come miglior candidato. L'instabilità sequenziale per i vari criteri di selezione è calcolata con $L = 70$ e vale:

```
data.frame(matrix(round(instab.assess(dati, 70, c(3,2,3), perc =
F),3),1,3,
       dimnames=list("k", c("AIC","BIC","HQ"))))
```

	AIC	BIC	HQ
k	0.8	1	0.957

Tutti i criteri sono abbastanza stabili. Non dovrebbe esserci un sostanziale miglioramento della performance usando l'algoritmo AFTER in questo contesto. L'$ASEP(138,208)$ calcolato utilizzando i diversi algoritmi è riportato sotto:

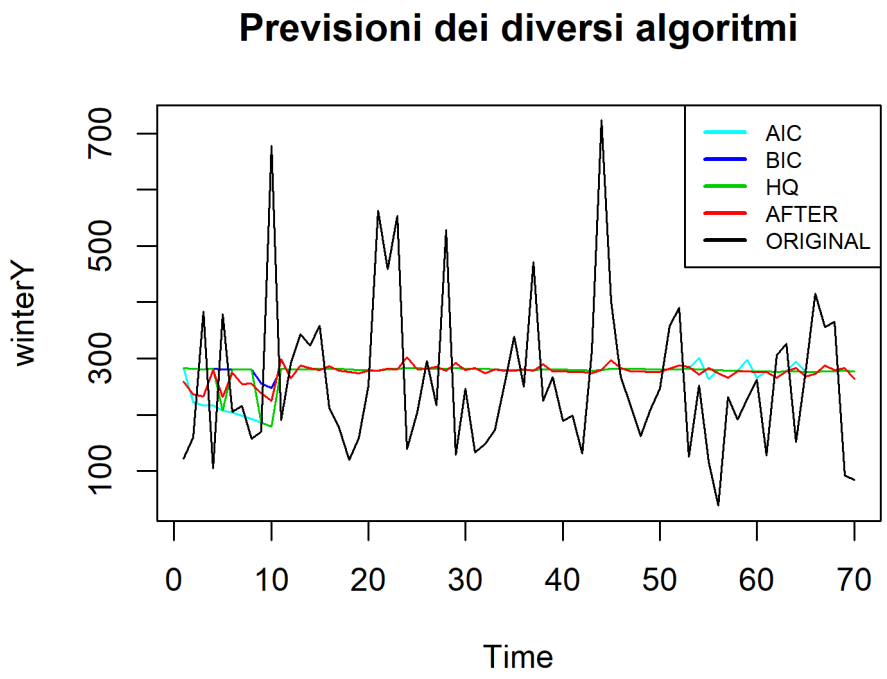
```
foo <- ICmodels(dati, 70, order=c(3,2,3))
foo2 <- AFTER(dati, 70, c(3,2,3))

winterY <- cbind(foo$yhat, foo2$yhat, foo$truth)
winterY <- ts(winterY, names=c("AIC","BIC","HQ","AFTER","ORIGINAL
"))

x <- c(foo$asep, foo2$asep)
winter <- data.frame(matrix(x, 1, 4, byrow=T,
                  dimnames = list(c("ASEP"), c("AIC","BIC","HQ","
AFTER"))))
winter
```

	AIC	BIC	HQ	AFTER
ASEP	19887.24	19267.78	20360.87	19455.45

```
plot(winterY, plot.type = "single", col=5:1,
     main="Previsioni dei diversi algoritmi")
legend("topright", legend=colnames(winterY), lwd=2, col=5:1, cex=
0.7)
```



Anche in questo contesto in cui l'algoritmo AFTER avrebbe dovuto performare peggio, comunque si avvicina molto al migliore errore di previsione tenuto dal BIC (il più stabile dei tre criteri).

CONCLUSIONI

La scelta del modello migliore per la previsione può presentare vari problemi, proprio per tale motivo, in questo elaborato viene proposto la combinazione come metodo, e in particolare si introduce l'algoritmo AFTER.

L'idea alla base è che quando c'è molta incertezza nella scelta del miglior modello, utilizzando la combinazione si tende a ridurre l'instabilità e a migliorare l'accuratezza della previsione. Questo viene dimostrato sia dalle simulazioni che dagli esempi di applicazioni su dati reali.

Per la selezione dei modelli si sono presi in considerazione criteri per misurare la stabilità del processo, così da poter vedere quanto sia affidabile il modello selezionato e quindi anche la corrispondente previsione.

Se vi è un'apparente instabilità allora conviene utilizzare la combinazione come metodo alternativo.

Si riassumono qui ora i risultati ricavati dalle simulazioni e applicazioni a dati reali con i modelli ARIMA:

1. La selezione del modello è preferibile quando c'è poca instabilità, quindi abbiamo poche difficoltà a trovare il modello migliore.
2. Se abbiamo un'incertezza nella scelta non trascurabile allora AFTER tende ad avere una performance migliore.
3. La misura di instabilità proposta è sensibile agli indicatori di incertezza nella selezione del modello e quindi può fornirci informazioni utili nella valutazione delle previsioni generate con questo metodo.

APPENDICE A

R e CODICE

R è un linguaggio di programmazione creato per gestire l'analisi statistica e la rappresentazione grafica dei suoi *output*.

La parte seguente rappresenta il codice usato per l'analisi dei dati reali riportati nel capitolo 6.

Funzione che sceglie il modello migliore basandosi su AIC, BIC e HQ:

```
choosing.best <- function(dati, P, D, Q, perc=F)
{
  foo <- as.matrix(expand.grid(0:P, 0:D, 0:Q))
  foo <- cbind(foo, matrix(NA, prod(c(P, D, Q)+1), 3))
  colnames(foo) <- c("p","d","q","AIC","BIC","HQ")
  n <- length(dati); err <- 0
  for(j in 1:prod(c(P, D, Q)+1))
  {
    m <- try(arima(dati, foo[j,1:3], optim.control = list(maxit=1500),
               method = "ML", transform.pars = T), silent = T)
    if(class(m)=="try-error")
    {
      err = err+1
      foo[j,4:6] <- Inf
    }
    else
    {
      foo[j,4] <- -2*m$loglik+2*(foo[j,1]+foo[j,3])
      foo[j,5] <- -2*m$loglik+(foo[j,1]+foo[j,3])*log(n)
      foo[j,6] <- -2*m$loglik+2*(foo[j,1]+foo[j,3])*log(log(n))
    }
    if(perc)
      cat(round(j/prod(c(P, Q, D)+1),2), " ")
  }
  if(perc)
    cat("\n\n")
  best <- matrix(NA, 3, 3, dimnames=
                   list(IC=c("AIC","BIC","HQ"), order=c("p","d","q")))
  best[1,] <- foo[which.min(foo[,4]), 1:3]
  best[2,] <- foo[which.min(foo[,5]), 1:3]
  best[3,] <- foo[which.min(foo[,6]), 1:3]

  return(best)
}
```

Legenda degli argomenti:

- dati: serie storica considerata;
- P, D, Q: ordine massimo dei modelli ARIMA considerati; più precisamente, come candidati per scegliere il modello migliore verranno considerati i modelli $ARIMA(p,d,q), \quad p = 0 \dots P, q = 0 \dots Q, d = 0 \dots D$;
- perc: Logico; consente di visualizzare l'avanzamento dell'algoritmo nella ricerca.

Per calcolare l'instabilità sequenziale dei vari indici è stato implementato il seguente algoritmo:

```
instab.assess <- function(dati, L, order, standard=NULL, perc=F)
{
  n <- length(dati)
  start <- n-L; end <- n-1
  if(is.null(standard))
    standard <- choosing.best(dati, order[1], order[2], order[3])
  instability <- numeric(3); attr(instability, "names") <- c("AIC","BI
C","HQ")

  for(j in start:end)
  {
    foo <- choosing.best(ts(dati[1:j]), order[1], order[2], order[3])
    instability <- instability + as.numeric(apply(foo==standard, 1, su
m)==3)
    if(perc)
    {
      cat(round((j-start+1)/L, 2), " ", "\n")
      print(foo); cat("\n\n")
    }

  }
  if(perc)
    cat("\n")

  return(instability/L)
}
```

Legenda degli argomenti:

- dati: serie storica considerata;
- order: vettore di lunghezza 3 del tipo (P, D, Q), dove P, D, Q assumono il significato descritto sopra. Questa variabile sostituisce e compatta a tutti gli effetti le 3 variabili sopra citate;

- L: Scalare che definisce quante osservazioni togliere dalla testa della serie; più precisamente, dato L, verranno considerate le serie storiche $\{y\}_1^j, \quad j = n - L + 1, \dots, n - 1$ per calcolare l'instabilità sequenziale;
- standard: opzionale, una matrice 3×3 che contenga, per riga, l'ordine del modello ARIMA selezionato da AIC, BIC e HQ rispettivamente; se non definito l'algoritmo richiama la funzione choosing.best sopra descritta per ricavare tale matrice;
- perc: Logico; consente di visualizzare l'avanzamento dell'algoritmo nella ricerca.

La funzione che calcola le previsioni utilizzando il criterio classico di selezione è la seguente:

```
ICmodels <- function(dati, n0, order, standard = NULL, perc=F)
{
  n <- length(dati); se <- rep(NA, 3)
  yhat <- matrix(NA, n0, 3, dimnames=list(t = (n-n0+1):n, c("AIC","BIC
","HQ")))

  for(j in (n-n0):(n-1))
  {
    standard <- choosing.best(dati[1:j], order[1],order[2],order[3])
    i <- j-n+n0+1
    arimaAIC <- arima(ts(dati[1:j]), standard[1,], optim.control = lis
t(maxit=1500))
    arimaBIC <- arima(ts(dati[1:j]), standard[2,], optim.control = lis
t(maxit=1500))
    arimaHQ <- arima(ts(dati[1:j]), standard[3,], optim.control = list
(maxit=1500))

    foo <- predict(arimaAIC, 1, se.fit = T)
    yhat[i,1] <- foo$pred; se[1] <- foo$se
    foo <- predict(arimaBIC, 1, se.fit = T)
    yhat[i,2] <- foo$pred; se[2] <- foo$se
    foo <- predict(arimaHQ, 1, se.fit = T)
    yhat[i,3] <- foo$pred; se[3] <- foo$se

    if(perc)
    {
      cat(round(i/n0, 2), "\n")
      print(standard)
      cat("\n\n")
    }
  }

  return(list(yhat=yhat,
              asep = apply(yhat-dati[(n-n0+1):n], 2, function(x) sum(x
^2)/n0),
```

```
              truth=dati[(n-n0+1):n]))
}
```

Legenda degli argomenti:

- dati, order, standard e perc: vedi sopra;
- n_0: Numero di osservazioni, in testa alla serie, utilizzate per verificare la bontà delle previsioni.

L'algoritmo AFTER è il seguente:

```
AFTER <- function(dati, n0, order, perc=F)
{
  # Initialising parameters
  n <- length(dati); truth <- dati[(n-n0+1):n]
  pred <- se <- W <- matrix(NA, prod(order+1), n0,
                            dimnames = list(models = 1:(prod(order+1)),
                                            period = (n-n0+1):n))
  start <- n-n0; end <- n-1

  # Estimating models and forecasting
  for(t in start:end)
  {
    j <- 0
    for(d in 0:order[2])
    {
      for(p in 0:order[1])
      {
        for(q in 0:order[3])
        {
          try(m <- arima(dati[1:t], c(p,d,q), optim.control = list(max
it=1500)), silent = T)
          j <- j+1
          yhat <- predict(m, 1, se.fit = T)
          pred[j,t-n+n0+1] <- yhat$pred; se[j,t-n+n0+1] <- yhat$se
        }
      }
    }
    if(perc)
      cat(round((t-n+n0+1)/n0, 2), " ")
  }

  # Weights
  W[,1] <- 1/nrow(W)
  for(i in 2:ncol(W))
  {
    W[,i] <- W[,i-1]/se[,i-1]*exp(-((truth[i-1]-pred[,i-1])^2/(2*se[,i
-1]^2)))/
      (sum(W[,i-1]/se[,i-1]*exp(-((truth[i-1]-pred[,i-1])^2/(2*se[,i-1]
^2)))))
  }
```

```
  # Predictions
  yhat <- apply(W*pred, 2, sum)
  return(list(yhat = yhat, asep = sum((yhat-truth)^2)/n0, w = W, truth
=truth))
}
```

Gli argomenti sono stati definiti come sopra.

BIBLIOGRAFIA

Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society*, *20*(4), 451-468.

Chatfield, C. (2000). *Time-series forecasting*. CRC Press.

Di Fonzo, T., & Lisi, F. (2007). *Serie storiche economiche*. Carocci.

Granger, C. W. (1989). Invited review combining forecasts—twenty years later. *Journal of Forecasting*, *8*(3), 167-173.

Hipel, K. W., & McLeod, A. I. (1994). *Time series modelling of water resources and environmental systems* (Vol. 45). Elsevier.

Makridakis, S., Wheelwright, S., & Hyndman, F. (1998). Methods and applications.

Muggeo, V. M. (2002). Il linguaggio R: concetti introduttivi ed esempi. *Published on the URL: http://www. cran. r-project. org/doc/contrib/nozioniR. pdf*.

Wallis, K. F. (2011). Combining forecasts–forty years later. *Applied Financial Economics*, *21*(1-2), 33-41.

Wei, G. C., & Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American statistical Association*, *85*(411), 699-704.

Yang, Y. (2003). Regression with multiple candidate models: selecting or mixing?. *Statistica Sinica*, 783-809.

Yang, Y. (2004). Combining forecasting procedures: some theoretical results. *Econometric Theory*, *20*(1), 176-222.

Zou, H., & Yang, Y. (2004). Combining time series models for forecasting. *International journal of Forecasting*, *20*(1), 69-84.