

A.M.A.S.E.S.
ASSOCIAZIONE PER LA MATEMATICA APPLICATA
ALLE SCIENZE ECONOMICHE E SOCIALI

ATTI
DEL VENTESIMO
CONVEGNO ANNUALE
A.M.A.S.E.S.

Urbino, 5-7 settembre 1996

Università degli Studi di Urbino
Con il Contributo del Consiglio Nazionale delle Ricerche

MODELLI PREVISIVI NEURALI: UN'APPLICAZIONE AL MERCATO FINANZIARIO ITALIANO

PierLuigi BELCARO, Elio CANESTRELLI e Marco CORAZZA

Dip. di Matematica Applicata ed Informatica - Università di Venezia

1. INTRODUZIONE

In letteratura è presente un crescente interesse in relazione alla predicibilità dei mercati finanziari. L'ipotesi classica, secondo cui le variazioni dei prezzi delle attività finanziarie sono assunte indipendenti e log-normalmente distribuite, non è accettata in un crescente numero di lavori empirici. In particolare, le distribuzioni empiriche delle variazioni logaritmiche di tali prezzi risultano caratterizzate dalla presenza di un numero significativo di outliers, da non stazionarietà nel livello della varianza, da asimmetria (ad esempio si veda [20], [9], [10], [26] e [6]) e da dipendenza nel breve e nel lungo periodo (ad esempio si veda [12], [14], [17], [21], [22], [6] e [5]). Sulla base di ciò, alcuni autori congetturano che queste variazioni possano essere dipendenti in maniera altamente non lineare (ad esempio si veda [25], [15], [16], [21], [3], [18] e [22]).

In lavori precedenti alcuni autori hanno analizzato il mercato finanziario italiano, riscontrando tracce evidenti di predicibilità [7], [4], [5]. Nel presente lavoro ci proponiamo di riprendere la medesima linea di ricerca. L'oggetto dell'analisi sono i rendimenti logaritmici giornalieri dell'indice della borsa italiana (COMIT) e lo strumento utilizzato è costituito da alcuni particolari modelli non lineari e non parametrici derivanti dal campo della soft Artificial Intelligence (sAI), le cosiddette Artificial Neural Networks (ANNs) a struttura (feedforward) Multi-Layer Perceptron (MLP). Ancora una volta i risultati che si ottengono rivelano tracce evidenti di predicibilità e sono connessi a prestazioni interessanti in relazione al trend ed al segno delle serie storiche prese in esame.

La parte rimanente del presente articolo è strutturata come segue. Nella sezione 2. forniamo sinteticamente alcuni aspetti teorici ed operativi preliminari; nella sezione 3. riportiamo l'esposizione dettagliata dell'approccio adottato; nella sezione 4. descriviamo i dati utilizzati; nella sezione 5. riferiamo i risultati ottenuti in relazione alla messa a punto di alcuni dei modelli MLP sviluppati nello studio; nella sezione 6. esponiamo i risultati dell'analisi empirica e nella sezione 7. riportiamo alcune considerazioni finali.

2. ASPETTI TEORICI ED OPERATIVI

In questa sezione forniamo in modo sintetico alcuni elementi richiesti per lo sviluppo del nostro studio.

2.1. Il modello feedforward Multi-Layer Perceptron (MLP)

I modelli feedforward MLP costituiscono semplici strumenti quantitativi, in grado di rappresentare, in maniera non parametrica, relazioni (non lineari) della forma $y = f(\mathbf{x}): A \subseteq \mathbb{R}^m \rightarrow B \subseteq \mathbb{R}^n$, dove A è lo spazio degli input e B è lo spazio degli output. Modelli di questo tipo presentano la struttura di grafi orientati ad archi pesati, i cui nodi siano disposti in uno strato di input (m nodi), in uno o in più strati nascosti ed in uno strato finale di output (n nodi). Ciascuno strato risulta completamente connesso con gli strati ad esso immediatamente successivi, senza alcuna connessione tra unità appartenenti ad un medesimo strato. Grazie a questa morfologia, il segnale in input viene propagato in modo unidirezionale dallo strato di input, attraverso gli strati nascosti, fino allo strato finale di output. Ciascun nodo della rete, ad eccezione di quelli appartenenti allo strato di input, è in grado di eseguire operazioni di computazione e di trasformazione (o trasferimento). In particolare, due funzioni caratterizzano ognuno di questi nodi. La prima determina l'input $a_{l,k}$ del k -esimo nodo dell' l -esimo strato, con $l = 2, \dots, L$ e $k = 1, \dots, K_l$, calcolando una "aggregazione" pesata dei segnali confluenti nel nodo medesimo e la seconda determina l'output $\tau_{k,l+1,h}$ propagato dal k -esimo nodo dell' l -esimo strato all' h -esimo nodo dell' $(l+1)$ -esimo strato, con $h = 1, \dots, K_{l+1}$, trasformando l'input $a_{l,k}$ (per maggiori dettagli si veda [8] e [13]).

Naturalmente, in questa classe di modelli rivestono un ruolo essenziale i parametri liberi, cioè i pesi $w_{i,l,k}$, con $i = 1, \dots, I_{l-1} + 1$, $l = 2, \dots, L$ e $k = 1, \dots, K_l$, ognuno, rispettivamente, associato all'arco che connette l' i -esimo nodo dell' $(l-1)$ -esimo strato con il k -esimo nodo dell' l -esimo strato. Allo scopo di determinare i loro valori "ottimali", utilizziamo il seguente algoritmo.

2.2. L'algoritmo di Error Back Propagation (EBP)

I valori ottimali dei pesi associati ad un modello MLP sono determinati sulla base di una procedura estimativa iterativa, denominata algoritmo di addestramento supervisionato, alla cui base si pone la metodologia di Discesa Rapida del Gradiente (DRG). Nell'ambito di questi algoritmi, dopo un'operazione di inizializzazione casuale dei pesi, le coppie di input-output appartenenti ad un prefissato insieme di dati $D = \{(\mathbf{x}_p, \mathbf{y}_p), p = 1, \dots, P\}$ vengono iterativamente presentate al modello MLP fintantoché, attraverso opportune operazioni di modifica dei pesi, non si

raggiunga il minimo (assoluto) di una funzione di costo predefinita. Una tipica funzione di costo è la seguente:

$$(2.2.1) \quad E = \frac{1}{2P} \sum_{p=1}^P \sum_{j=1}^n \left[y_{p,j}^* (w_{1,1,1}, \dots, w_{I_{l-1}+1, L, K_l}) - y_{p,j} \right]^2$$

dove P è il numero delle coppie di input-output appartenenti al prefissato insieme di dati, n è la dimensione dello spazio di output, $y_{p,j}^* (w_{1,1,1}, \dots, w_{I_{l-1}+1, L, K_l})$ o, più semplicemente, $y_{p,j}^*$, con $j = 1, \dots, n$, sono gli elementi del p -esimo vettore degli output calcolati dal modello MLP (il cui valore dipende da quello assunto dai pesi di iterazione in iterazione) e $y_{p,j}$ sono, infine, gli elementi del p -esimo vettore degli output desiderati. In particolare, l'aggiornamento dei pesi si ottiene retropropagando il valore assunto dalla funzione di costo dallo strato di output, attraverso gli strati nascosti, fino allo strato iniziale di input, nel modo seguente:

$$(2.2.2) \quad w_{i,l,k}(t+1) = w_{i,l,k}(t) + \Delta w_{i,l,k}(t+1) = w_{i,l,k}(t) - \eta \frac{\partial E}{\partial w_{i,l,k}}(t+1) \Bigg|_{(x_1, y_1), \dots, (x_p, y_p)},$$

$i = 1, \dots, I_{l-1} + 1, l = L, \dots, 2, k = 1, \dots, K_l$ e $t = 0, \dots, T$,

dove $w_{i,l,k}(q)$ è il valore, alla q -esima iterazione (od epoca) dell'algoritmo, del peso, $\Delta w_{i,l,k}(q)$ è il valore, alla medesima iterazione, dell'aggiornamento del peso stesso, $\eta \in]0, 1]$ è un parametro denominato tasso di apprendimento, $\left(\frac{\partial E}{\partial w_{i,l,k}} \right)(q) \Big|_{(x_1, y_1), \dots, (x_p, y_p)}$ è il valore, sempre alla stessa iterazione, assunto dalla derivata parziale della funzione di costo rispetto al peso stesso e T , infine, è il numero di iterazioni richieste dall'algoritmo d'addestramento (che risulta implicitamente definito da un'opportuna regola di stop-learning fissata inizialmente).

In un tale algoritmo è rilevante il ruolo assunto dalla regola di stop-learning, la cui versione classica risulta soddisfatta a quella T -esima iterazione dell'algoritmo in cui il livello d'errore rilevato $E(T)$ non sia superiore al valore di una prefissata soglia suddetta \bar{E} . Tuttavia, ad una scelta di questo tipo può accompagnarsi un fenomeno sfavorevole denominato iper-apprendimento, consistente nella possibilità che tale algoritmo individui relazioni funzionali, tra i vettori di input \mathbf{x} ed i vettori di output \mathbf{y} , inesistenti. Allo scopo di evitare il manifestarsi di tali inconvenienti, alcuni autori suggeriscono di utilizzare un criterio di stop-learning basato sulla Metodologia

di Discesa Concorrente dell'errore (MDC) (ad esempio si veda [8] e [13]). In un approccio così definito, l'insieme iniziale di dati D viene, prima di tutto, convenientemente suddiviso in due sottoinsiemi ad intersezione vuota, il *training set* (D_T) ed il *testing* (o *validation*) *set* (D_V) e, successivamente, l'algoritmo di apprendimento viene applicato iterativamente al training set D_T , finché non sia raggiunto il minimo (assoluto) della funzione di costo calcolata sul testing set D_V .

3. L'APPROCCIO SEGUITO PER L'ANALISI DELLA PREDICIBILITA'

In questa sezione vengono fornite in dettaglio le informazioni relative all'approccio che abbiamo sviluppato.

In generale, nell'ambito dello sviluppo di modelli MLP per la previsione, vari sono gli aspetti a cui si dovrebbe prestare attenzione e che potrebbero essere raggruppati nelle tre seguenti classi: una prima che considera la selezione delle variabili di input ed output e le operazioni di pre- e post-processamento dei dati [1], [23]; una seconda che esamina l'aspetto del design architetturale del modello; una terza che prevede l'analisi delle procedure di training e testing [1], [2], [23].

E' doveroso sottolineare che l'obbiettivo principale del nostro lavoro non è quello di mettere a punto strumenti di previsione puntuale particolarmente accurati quanto quello di sviluppare modelli che possano permetterci di cogliere indicazioni e segnali della presenza di predicibilità. In relazione al perseguimento di tale obbiettivo, abbiamo deciso di privilegiare alcuni degli aspetti tecnici sopra tratteggiati rispetto ad altri. In particolare, è possibile identificare i tre seguenti livelli di "attenzione": un primo livello di attenzione essenziale, un secondo livello di attenzione standard ed un terzo livello di attenzione elevata.

Si è dedicata attenzione a livello essenziale alle operazioni di selezione delle variabili di input ed output. Infatti, si utilizza un semplice approccio univariato, senza tuttavia perdere la consapevolezza che una scelta di questo tipo potrebbe risultare associata a situazioni di dissipazione informativa. Anche il livello di attenzione prestato alle operazioni di processamento dei dati è essenziale, poiché si utilizzano i dati non processati, lasciando, così, che i modelli sopportino tutto il carico di difficoltà di ordine computazionale derivanti da una tale scelta (per maggiori dettagli si veda [23]).

In secondo luogo, si presta un livello di attenzione standard alle operazioni di selezione delle funzioni di trasferimento, in conformità alle principali linee guida presenti in letteratura [1], [2], [23]. In particolare, per ogni nodo-processore dei nostri modelli MLP utilizziamo la classica funzione logistica

(per maggiori dettagli si veda [8] e [13]). E' da notare che alcuni autori, in relazione all'aggiornamento dei pesi nel corso della fase di training, suggeriscono di utilizzare la seguente versione modificata della regola di apprendimento supervisionato:

$$(3.1) \quad w_{i,l,k}(t+1) = (1+\beta)w_{i,l,k}(t) - \beta w_{i,l,k}(t-1) - (1-\beta)\eta \frac{\partial E}{\partial w_{i,l,k}}(t+1) \Bigg|_{(x_1, y_1), \dots, (x_p, y_p)}$$

$i = 1, \dots, I_{l-1} + 1, l = L, \dots, 2, k = 1, \dots, K_l$ e $t = 1, \dots, T,$

dove $\beta \in [0,1]$ è un parametro denominato momentum, in base al quale l'addestramento può risultare migliore, grazie all'introduzione di una sorta di memoria dei valori assunti dai pesi al termine delle precedenti iterazioni dell'algoritmo. Un livello standard di attenzione si presta, alla gestione tanto del tasso di apprendimento, quanto del momentum. In accordo alle metodologie classiche (ad esempio si veda [1] e [23]), il tasso di apprendimento dovrebbe subire opportune azioni continue di aggiustamento nel corso del training stage. Noi impieghiamo una regola di riduzione lineare a tratti, avviando l'addestramento dei modelli sulla base di un valore $\eta = 1$, che viene poi ridotto a mano a mano che l'apprendimento procede in modo efficace. Anche per ciò che riguarda la gestione del momentum, ci si conforma ai suggerimenti rintracciabili in buona parte della letteratura (ad esempio si veda [2] e [23]), avviando l'attività sperimentale sulla base di un valore $\beta = 0.9$. Anche le operazioni di inizializzazione ponderale e di scrambling dei dati ricevono un livello di attenzione standard. In particolare, alcuni autori propongono di addestrare una rete MLP sulla base di una serie di diverse configurazioni ponderali iniziali (ad esempio si veda [1] e [23]), cosicché sia possibile procedere allo sviluppo di più training stage, i cui esiti possano venir tra loro confrontati, allo scopo di prevenire l'insorgere di difficoltà, quali, ad esempio, il raggiungimento di un minimo relativo, anziché uno assoluto, della funzione di costo considerata. Anche lo scrambling dei dati, inoltre, è una proposta che viene presentata comunemente in letteratura, allo scopo di prevenire il verificarsi di problemi derivanti da uno sviluppo scorretto del training, quali, ad esempio, l'apprendimento di trend funzionali parassiti. In linea con tutto ciò, si adottano qui procedure di inizializzazione ponderale multipla e di scrambling dei dati.

Infine, è stato riservato un livello di attenzione elevata ai due seguenti aspetti: il design architetturale dell'MLP e la regola di stop-learning, dal momento che entrambi sembrano porsi come elementi essenziali [1], [13], [23]. Per ciò che riguarda il primo aspetto (la struttura architetturale dell'MLP) consideriamo

dapprima il numero di nodi presenti nello strato di input ed in quello di output e, successivamente, il dimensionamento della regione nascosta, operazione, quest'ultima, che costituisce uno dei punti critici della progettazione di una rete MLP. In particolare, allo scopo di individuare la lunghezza "ottimale" della finestra temporale di input, si prendono in considerazione due differenti valori del numero di nodi presenti all'interno dello strato di input, $m \in \{15, 22\}$, corrispondenti, rispettivamente, a 15 giorni (1 nodo di input per ciascun giorno compreso nella finestra temporale), cioè a circa tre settimane borsistiche (modelli di classe S, da *Shrunk*), ed a 22 giorni, cioè a circa un mese borsistico (modelli di classe W, da *Wide*). Oltre a ciò, allo scopo di monitorare l'evoluzione del mercato, si prendono in esame 4 differenti valori del numero di nodi presenti nello strato di output, $n \in \{1, 2, 3, 5\}$, risultando ciascun nodo di output associato alla corrispondente giornata di previsione immediatamente successiva all'ultima giornata compresa nella finestra di input. Quanto al dimensionamento della regione nascosta, si sviluppano tre differenti soluzioni architetture. Concordemente alle ragioni della loro comparsa e del loro sviluppo, queste tre impostazioni vengono qui denominate come segue: approccio Kolmogorov, approccio DARPA ed approccio Bottleneck.

Le reti MLP Kolmogorov vanno considerate congiuntamente al seguente teorema di esistenza, pubblicato nel 1957 dal matematico russo A. N. Kolmogorov e qui di seguito riportato come in [13]: "*Given any continuous function $f: [0, 1]^m \rightarrow \mathbb{R}^n$, $f(\mathbf{x}) = \mathbf{y}$, $f(\cdot)$ can be implemented exactly by a 3-layer feedforward neural network having m fanout processing elements in the first (\mathbf{x} input) layer, $(2m + 1)$ processing elements in the middle layer and n processing elements in the top (\mathbf{y} output) layer*". Le reti MLP di Kolmogorov forniscono generalmente risultati soddisfacenti nell'ambito dei problemi di previsione finanziaria (ad esempio si veda [1] e [23]), ma non va dimenticato che si tratta di un teorema di esistenza. Inoltre, solo la classe delle funzioni continue può essere presa in considerazione.

La soluzione architetture DARPA (la denominazione va riferita all'organismo statunitense *Defense Advanced Research Project Agency*) ha per oggetto reti MLP il cui numero di strati nascosti può essere superiore ad uno, in dipendenza dal carattere topologico di separabilità o non separabilità lineare attribuibile ai sottospazi vettoriali nei quali il dominio delle relazioni funzionali analizzate potrebbe venir suddiviso [2], [13], [23]. È stato teoricamente dimostrato che solo reti MLP provviste di due o più strati nascosti sono in grado di discriminare morfologie funzionali a carattere di non separabilità lineare tanto convessa quanto non convessa [13]. Allo scopo di assicurare potenzialità discriminanti, si prendono in esame qui reti MLP

DARPA dotate di due strati nascosti e, simultaneamente, si considerano differenti valori del numero di nodi che li compongono.

L'ultimo approccio architetturale sviluppato è rappresentato dai modelli Bottleneck, i quali risultano costituiti da reti MLP dotate di un solo strato nascosto il cui numero di nodi sia, in vario grado, ridotto rispetto al numero dei nodi presenti nello strato di input. Le reti MLP Bottleneck vengono qui prese in considerazione in ragione della sostanziale "robustezza" che emerge dai loro utilizzi empirici nel campo dell'analisi previsionale finanziaria [23]. In particolare, per individuare in quale misura influisca il sottodimensionamento dello strato nascosto, si mettono a punto più reti Bottleneck, dotate di numerosità dello strato nascosto progressivamente decrescente.

In relazione al secondo aspetto rilevante del lavoro, la regola di stop-learning, allo scopo di conseguire snellezza operativa, si adotta qui una versione opportunamente modificata della regola di stopping MDC (si veda la sottosezione 2.2). In particolare, in questa specifica versione, anziché monitorare la dinamica delle funzioni di costo nel corso del training stage, l'algoritmo di apprendimento (basato sulla metodologia EBP) viene iterato sul training set D_T fintantoché il massimo (assoluto) di una prefissata funzione di guadagno sul testing set D_V non sia raggiunto. La funzione di guadagno fornisce una sorta di misura dell'"abilità generalizzante" raggiunta dal modello, e può avere la forma seguente:

$$(3.2) \quad G(q) = \left[100 \frac{\sum_{p=1}^P s_p(q)}{P_V} \right] \% \in [0, 100], \quad s_p(q) = \begin{cases} 1 & \text{se } \sum_{j=1}^n (y_{p,j}^* - y_{p,j})^2 \leq \bar{E}_V \\ 0 & \text{altrimenti} \end{cases}$$

$$q = 1, \dots, T,$$

dove $G(q)$ rappresenta il valore, al termine della q -esima iterazione dell'algoritmo di apprendimento, della funzione di guadagno, $y_{p,j}^*(q)$, con $j = 1, \dots, n$, sono i valori, alla medesima iterazione, degli n elementi del p -esimo vettore di output calcolato dal modello e \bar{E}_V rappresenta un prefissato valore di soglia.

Riassumendo, considerando tutti i modelli di rete generati, si perviene ad un numero finale di realizzazioni modellistiche che sfiora le 600 unità.

4. IDATI

I dati che utilizzati sono costituiti dai rendimenti logaritmici calcolati sulla base delle quotazioni di chiusura giornaliera (cioè $r(t+1) = \ln[P(t+1)] - \ln[P(t)] = \ln[P(t+1)/P(t)]$) dell'indice generale

della borsa italiana COMIT. In particolare, si prendono in esame il periodo campione dal 4 Gennaio 1993 al 18 Ottobre 1994, per quanto riguarda le coppie input-output appartenenti al training set D_T e il periodo campione dal 19 Ottobre 1994 al 30 Dicembre 1994, per quanto riguarda le coppie input-output appartenenti al testing set D_V . Nel corso della fase previsionale viene preso in considerazione il periodo campione che va dal 2 Gennaio 1995 al 12 Aprile 1995, per quanto riguarda le coppie input-output appartenenti al forecasting (o running) set D_R . Riassumendo, l'insieme di dati iniziale delle coppie input-output risulta così suddiviso nei tre seguenti sottoinsiemi:

$$D_T = \{(\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_{P_T}, \mathbf{y}_{P_T})\}, \quad D_V = \{(\mathbf{x}_{P_T+1}, \mathbf{y}_{P_T+1}), \dots, (\mathbf{x}_{P_T+P_V}, \mathbf{y}_{P_T+P_V})\},$$

$$D_R = \{(\mathbf{x}_{P_T+P_V+1}, \mathbf{y}_{P_T+P_V+1}), \dots, (\mathbf{x}_{P_T+P_V+P_R}, \mathbf{y}_{P_T+P_V+P_R})\},$$

dove P_T , P_V e P_R sono i numeri delle coppie di input-output che appartengono, rispettivamente, a D_T , D_V e D_R , $\mathbf{x}_i = (r((i-1)(m+n)+1), \dots, r(im+(i-1)n))$ e $\mathbf{y}_i = (r(im+(i-1)n+1), \dots, r(i(m+n)))$, con $i = 1, \dots, P_T + P_V + P_R$.

Occorre qui sottolineare che il primo vettore di input ($\mathbf{x}_{P_T+P_V+1}$) del running set D_R risulta composto di elementi che non appartengono anche agli ultimi vettori del testing set D_V . Questa scelta comporta necessariamente la "perdita" di una parte delle previsioni collocate nello spazio di tempo immediatamente successivo rispetto all'orizzonte temporale di testing. In ragione della loro maggior vicinanza agli insiemi di training e testing, le previsioni che risultano così escluse potrebbero proprio essere le migliori.

5. I RISULTATI DELLE FASI DI TRAINING E DI TESTING

Questa sezione è dedicata all'esposizione dei risultati ottenuti nelle fasi di training e di testing.

(5.A) E' possibile rilevare che, in generale, l'impiego nel corso della fase di addestramento di un valore di momentum sembra non solo non produrre alcun effetto favorevole, ma, anzi, contribuire al "disorientamento" dell'intero processo di apprendimento. Conseguentemente, si addestra il complesso dei modelli di rete senza impiegare alcun momentum (cioè, $\beta = 0$).

(5.B) E' altresì possibile osservare che differenti inizializzazioni ponderali ed operazioni di scrambling dei dati sembrano produrre pochi o addirittura nessun effetto sulle prestazioni finali dei modelli.

(5.C) Relativamente ai modelli della classe Bottleneck, quelli che dimostrano di raggiungere le migliori prestazioni sono quelli il cui numero di nodi nello strato nascosto è pari ad un terzo della corrispondente grandezza relativa allo strato di input. Premesso ciò, tutte e tre le classi di reti MLP

(Kolmogorov, DARPA e Bottleneck) impegnate nelle fasi di training e testing manifestano livelli prestazionali complessivi sostanzialmente soddisfacenti.

(5.D) I modelli appartenenti alla classe DARPA fanno registrare, solo nei casi migliori, livelli di prestazione confrontabili con quelli rilevabili nelle altre classi. In altri termini, non sembra che l'analisi del problema richieda un livello elevato di potenzialità discriminante in relazione alle caratteristiche topologiche (si veda la sezione 3).

(5.E) I modelli di rete della classe W dimostrano, generalmente, livelli prestazionali più elevati rispetto ai modelli corrispondenti della classe S, indipendentemente dalla classe architettonica cui appartengono. Per ciò si potrebbe congetturare la presenza di un meccanismo di memoria a breve-medio termine, la cui azione metta in grado le reti di sfruttare vantaggiosamente l'informazione presente nelle determinazioni passate della serie storica.

(5.F) Infine, i migliori livelli prestazionali globali risultano associati alle reti della classe Bottleneck. Non è semplice comprendere l'eziologia fenomenologica che sottende risultati di questo tipo [13]. La congettura che si potrebbe avanzare riguarda la possibilità che strutture interpretative "essenziali" (quali sono le reti Bottleneck) riflettano un grado di analogia "essenzialità", seppur in senso lato, delle sottostanti relazioni funzionali.

Nella tabella seguente si riportano i valori massimi assunti dalla funzione di guadagno sul testing set D_T e sul training set D_T . Le prestazioni delle reti della classe DARPA non compaiono per le ragioni sopra esposte.

Tabella 5.1 - Valori medi (classi W ed S)

Modellistica MLP	Valore massimo della funzione di guadagno (Testing set)	Valore corrispondente della funzione di guadagno (Training set)
Classe W-Kolmogorov	88.33%	82.33%
Classe W-Bottleneck	93.33%	84.33%
Classe S-Kolmogorov	89.00%	81.67%
Classe S-Bottleneck	91.33%	83.67%

6. I RISULTATI DELLA FASE PREVISIONALE

In questa sezione esponiamo i risultati ottenuti nella fase di previsione. Il primo problema che si pone è quello di mettere a punto una metodologia idonea alla rappresentazione ed alla valutazione della "distanza" che separa i vettori degli output calcolati y_p^* dai corrispondenti vettori dei valori desiderati y_p , con $p = 1, \dots, P_R$, dove P_R è il numero di coppie di input-output impegnate nella fase previsionale. Un approccio potrebbe essere quello di considerare, inizialmente, differenti indicatori prestazionali e di proporre,

successivamente, sulla loro base, un giudizio finale a carattere sintetico. In relazione all'adozione di una tale soluzione, varie sono le proposte ed i suggerimenti presenti nella recente letteratura (per maggiori dettagli si veda [23] e [24]). In questa sede si ricorre ad un sistema indicatoriale opportunamente sviluppato, all'interno del quale confluiscono indicatori espressamente provenienti dalla statistica "classica" insieme ad altri indicatori realizzati *ad hoc*. In particolare, allo scopo di esaminare il loro sviluppo dinamico a mano a mano che il numero di vettori di output calcolati aumenta, vengono calcolati per tutti gli $S = S_1, \dots, P_R$, con $S_i \in \{1, 2\}$. Gli indicatori realizzati *ad hoc* sono:

$$(6.1) \quad CT_{S,h} = \left(\frac{100}{S-1} \sum_{i=1}^{S-1} \alpha_i \right) \% \in [0, 100], \quad \alpha_i = \begin{cases} 1 & \text{se } (y_{i+1,h} - y_{i,h})(y_{i+1,h}^* - y_{i,h}) > 0 \\ 1 & \text{se } y_{i+1,h} = y_{i,h} = y_{i+1,h}^* \\ 0 & \text{altrimenti} \end{cases}$$

$$(6.2) \quad CS_{S,h} = \left(\frac{100}{S} \sum_{i=1}^S \beta_i \right) \% \in [0, 100], \quad \beta_i = \begin{cases} 1 & \text{se } y_{i,h} y_{i,h}^* > 0 \\ 1 & \text{se } y_{i,h} = y_{i,h}^* = 0 \\ 0 & \text{altrimenti} \end{cases}$$

$$(6.3) \quad ANP_{S,h} = \sum_{i=1}^S \gamma_i y_{i,h} \in]-\infty, +\infty[, \quad \gamma_i = \begin{cases} 1 & \text{se } y_{i,h}^* > 0 \\ 0 & \text{se } y_{i,h}^* = 0 \\ -1 & \text{se } y_{i,h}^* < 0 \end{cases}$$

$$(6.4) \quad DARNP_{S,h} = \left(\sum_{i=1}^S \gamma_i y_{i,h} \right) / \sum_{i=1}^S |y_{i,h}| \in [-100, 100], \quad \gamma_i = \begin{cases} 1 & \text{se } y_{i,h}^* > 0 \\ 0 & \text{se } y_{i,h}^* = 0 \\ -1 & \text{se } y_{i,h}^* < 0 \end{cases}$$

dove $y_{i,h}^*$, con $i = S_1, \dots, P_R$ e $h \in \{1, 2, 3, 5\}$ è l' h -esimo elemento (cioè, la previsione h steps ahead) dell' i -esimo vettore di output calcolati, mentre $y_{i,h}$ è l' h -esimo elemento dell' i -esimo vettore degli output desiderati.

Per ciò che riguarda gli indicatori CT (*Correct Trend*) e CS (*Correct Signum*), è]50%, 100%] l'intervallo ottimale. Generalmente, ci si aspetta che i valori dell'indicatore CT siano più elevati rispetto a quelli dell'indicatore CS, poiché quest'ultimo risulta più direttamente associato a previsioni puntuali accurate. Per l'indicatore ANP (*Absolute Net Profitability*) ed per il DARNP (*Daily Average Relative Profitability*) sono i valori superiori allo zero a costituire il livello ottimale minimo. L'ANP ed il DARNP rappresentano misure di profittabilità basate su di una semplice strategia di trading, secondo

la quale un'attività finanziaria dovrebbe essere acquistata oppure venduta a seconda della dinamica di ascesa-discesa del suo prezzo.

Va sottolineato che, allo scopo di estrarre una valutazione sintetica finale, si è attribuita maggiore rilevanza agli indicatori di trend, di segno e di profittabilità piuttosto che ad altri, questo poiché uno degli obiettivi del presente contributo è quello di dare maggior risalto agli aspetti operativi. Inoltre, ricordando ancora che l'intero complesso indicatoriale viene applicato, secondo le modalità di sviluppo dinamico cui si accennava sopra, a ciascuna colonna di output delle reti, a ciascun modello MLP risultano associate tante valutazioni globali quanti sono i nodi di cui è composto il suo strato di output. Tali valutazioni sono espresse sulla base di una scala di giudizio in cui A rappresenta il punteggio più elevato, mentre D quello più basso (si vedano le Figure dalla 6.1 alla 6.4).

Fig. 6.1 - Modelli MLP W-Kolmogorov

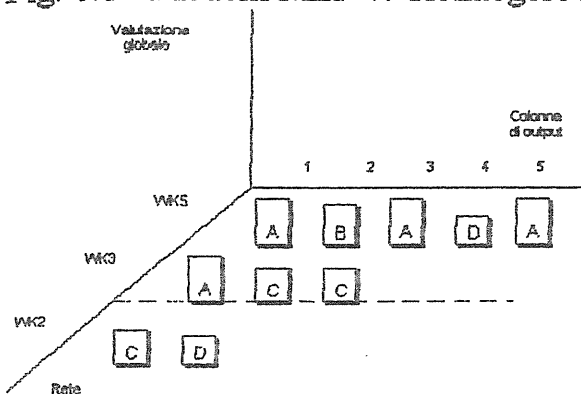


Fig. 6.2 - Modelli MLP S-Kolmogorov

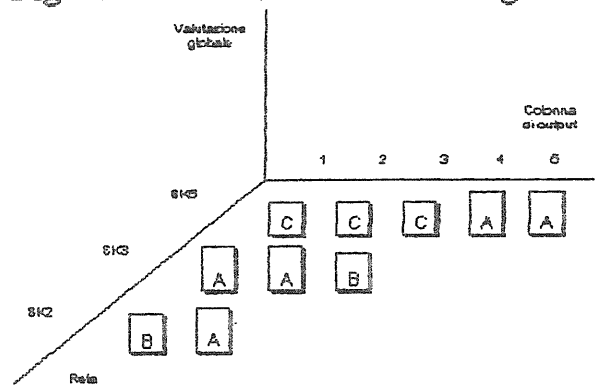


Fig. 6.3 - Modelli MLP W-Bottleneck

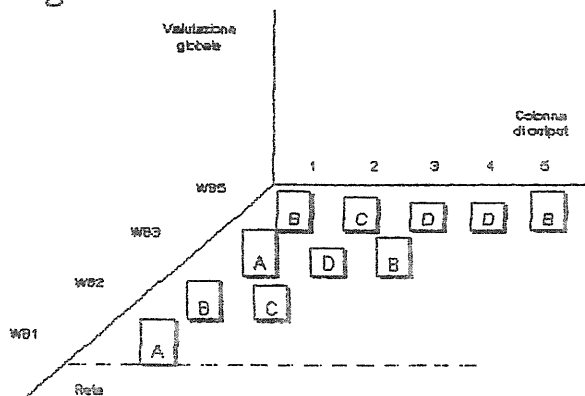
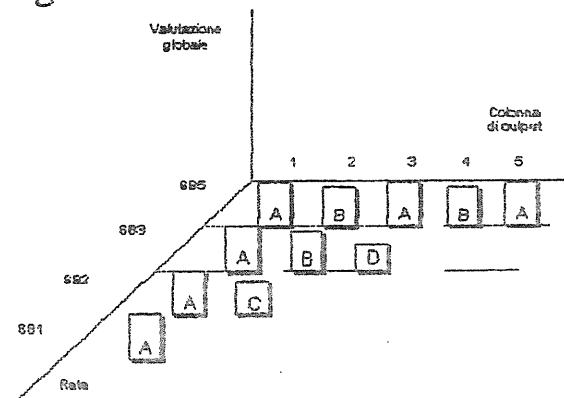


Fig. 6.4 - Modelli MLP S-Bottleneck



Occorre sottolineare che un'operazione valutativa come quella appena descritta configura comunque un processo di natura soggettiva, in cui sono gli indicatori CT e CS e dell'indicatore di profittabilità ANP (vs. la profittabilità

totale) a costituire la piattaforma segnaletica di sostegno dell'intero processo valutativo (si vedano le Figure dalla 6.5 alla 6.8).

Figura 6.5 - S-Bottleneck-5 (Colonna 1 - Valutazione globale: A)

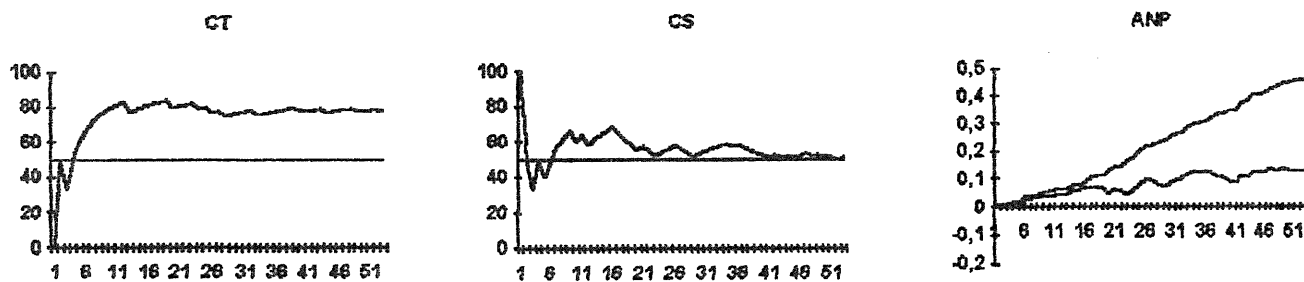


Figura 6.6 - W-Bottleneck-3 (Colonna 3 - Valutazione globale: B)

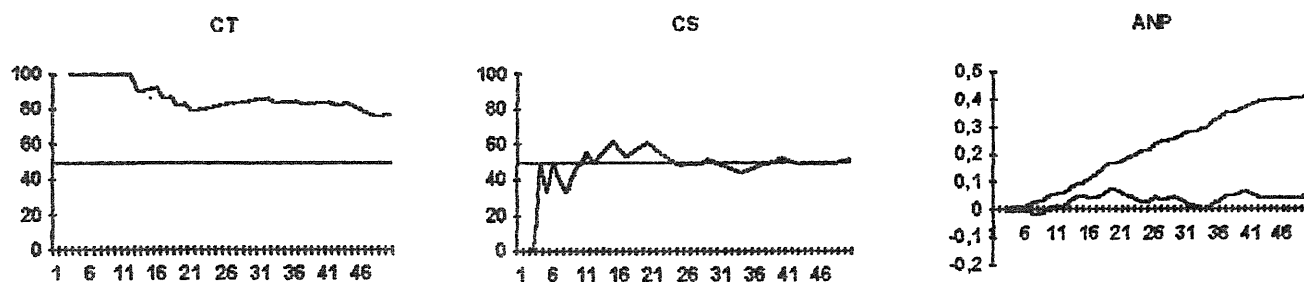


Figura 6.7 - S-Kolmogorov-5 (Colonna 3 - Valutazione globale: C)

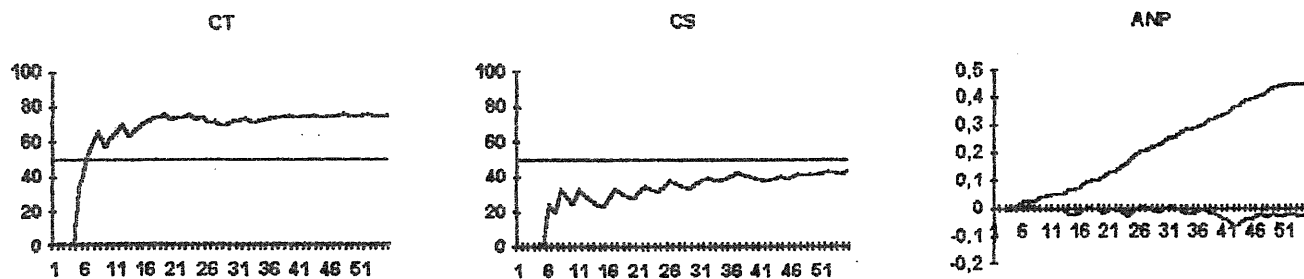
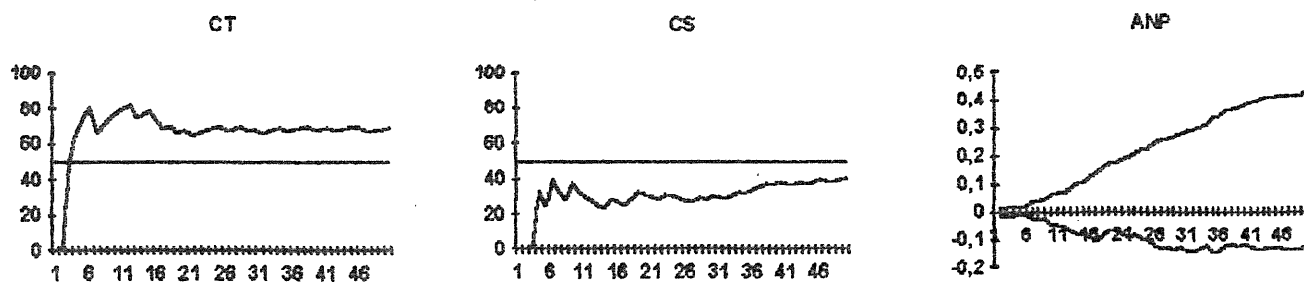


Figura 6.8 - W-Kolmogorov-2 (Colonna 2 - Valutazione globale: D)



Dall'analisi del complesso dei risultati ottenuti nella fase previsionale emergono alcune macro-caratteristiche comuni.

(6.A) La dimensione relativa degli strati di input (S vs. W), così come il dimensionamento delle finestre di output (1, 2, 3 o 5 nodi) non sembrano, a differenza di quanto accadeva nelle fasi di training e di testing, giocare un ruolo particolarmente significativo. Inoltre, appare meno rilevante anche la struttura architeturale (reti Kolmogorov vs. reti Bottleneck) se, anche in questo caso, manteniamo come termine di paragone, i risultati ottenuti nelle fasi di training e testing. Ciò potrebbe venir ricondotto ad una sensibilità dei modelli di rete (indipendentemente dalla classe architeturale a cui appartengano) alla distanza temporale che separa i dati contenuti negli insiemi di testing dai dati contenuti negli insiemi di running.

(6.B) La dimensione degli errori non appare trascurabile. Ciò va, d'altra parte, ricondotto alla decisione di sviluppare strumenti previsionali neurali "semplici". Inoltre, i valori del coefficiente di correlazione R appaiono, talvolta, poco elevati, anche se, sostanzialmente, si possono dire accettabili, avendosi, generalmente, $R \in [15\%, 25\%]$.

(6.C) In generale dei casi, i valori segnati dall'indicatore CT si collocano ad un livello soddisfacente, sintomo, quest'ultimo, di un efficace apprendimento da parte dei modelli delle dinamiche di ascesa e discesa. Infatti, per lo più si trova $CT \in [60\%, 80\%]$.

(6.D) Per ciò che riguarda i livelli registrati dall'indicatore CS, un numero ristretto di reti dimostrano manifestazioni dell'indice al di sotto della soglia del 50%, mentre un altro ridotto numero di modelli permette di osservare livelli intorno alla soglia del 50%. Tuttavia, la parte preponderante dei modelli è in grado di effettuare previsioni cui risulta associato un livello di CS sistematicamente superiore al 50%, trovandosi, nei casi migliori, $CS \in [60\%, 70\%]$. Sensibilmente dipendenti dai risultati relativi all'indicatore CS sono le prestazioni della profittabilità, cioè degli indicatori ANP e DARNP, poiché la strategia di trading che sostiene quest'ultime misure è basata proprio sulla dinamica del segno (si vedano le espressioni (6.5) e (6.8)). In connessione ai migliori livelli prestazionali di CS, è possibile individuare situazioni per cui $DARNP \in [70\%, 80\%]$.

(6.E) In riferimento ai valori dell'indicatore di Theil e dell'indicatore di varianza media relativa, è possibile rilevare la presenza di una sostanziale componente strutturale all'interno delle relazioni funzionali indagate, dal momento che entrambi i livelli indicatoriali si attestano intorno a valori soddisfacentemente bassi. Inoltre, da un confronto tra i livelli di entrambi gli indici, risulta che, talvolta, previsioni formulate sulla base di semplici medie

storiche progressive potrebbero competere con quelle prodotte dagli stessi modelli MLP.

(6.F) Per concludere, si nota che, in corrispondenza delle colonne centrali di previsione dei modelli MLP, si manifesta una sorta di fenomeno di flessione prestazionale.

Un ultimo aspetto va posto in rilievo. La metodologia dinamica di rilevazione dei livelli indicatoriali ci pone in grado di monitorare l'evoluzione di ciascun indicatore. Quando il monitoraggio dei livelli prestazionali di un opportuno sottoinsieme indicatoriale lasci apparire visibili segnali di declino, potrebbe risultare conveniente arrestare la fase previsionale, procedere ad un riaddestramento delle reti sulla base di training e testing set debitamente aggiornati e, dopo, riavviare le operazioni di previsione. In questo modo, la distanza cronologica che separa gli insiemi di training e testing da quello di running verrebbe a ridursi, lasciando spazio al ripristino di "opportunità previsive" più favorevoli.

7. CONSIDERAZIONI CONCLUSIVE

Questa sezione è dedicata alle seguenti considerazioni finali.

(7.A) Nonostante lo sviluppo della strumentazione MLP abbia seguito linee realizzative di essenzialità e di attenzione standard in relazione a determinati aspetti tecnici (si veda la sezione 3.), i risultati finali ottenuti rivelano evidenti tracce di predicibilità nel mercato finanziario italiano. Su questa base empirica appare ragionevole congetturare che, dedicando un livello di attenzione elevato anche agli aspetti testé delineati, potrebbe risultare possibile ottenere previsioni puntuali più accurate di quelle raggiunte nel presente lavoro.

(7.B) Gli evidenti segnali registrati della presenza di predicibilità conducono ad esprimere la congettura secondo cui, all'origine dei processi di generazione dei rendimenti, possano operare congiuntamente tanto meccanismi di natura stocastica quanto meccanismi non lineari di natura deterministica.

(7.C) Per concludere, date le suddette manifestazioni di predicibilità, sembra ormai necessario lo sviluppo di nuovi schemi teorici, all'interno dei quali le modellistiche classiche (quali, tipicamente, lo schema di Black e Scholes), che si fondano sull'assunzione della presenza di distribuzioni (log-normalmente) indipendenti ed identicamente distribuite, trovino occasione di ampliamento.

8. RIFERIMENTI BIBLIOGRAFICI

[1] AZOFF, E. M. (1994) *Neural Network Time Series Forecasting of Financial Markets*. John Wiley & Sons, New York.

- [2] BAESTAENS, D. E., VAN DEN BERGH, W. M. e WOOD, D. (1994) *Neural Network Solutions for Trading in Financial Markets*. Pitman Publishing, London.
- [3] BOOTH, G. G., MARTIKAINEN, T., SARKAR, S. S., VIRTANEN, I. e YLI-OLLI, P. (1992) Nonlinear Dependence in Finnish Stock Returns, lecture presentata all'*Euro Working Group on Financial Modelling*, 12th Meeting, Helsinki.
- [4] CORAZZA, M. (1993/94) *Caso e Caos Deterministico: un Approccio all'Analisi delle Leggi di Evoluzione dei Prezzi Speculativi*. Tesi di Dottorato di Ricerca, Università di Brescia, Brescia.
- [5] CORAZZA, M. (1996) Long-term Memory Stability in the Italian Stock Market, lecture presented at *Letters in Public Finance*, 1st Meeting, Acquafredda di Maratea (sottoposto a *Economic & Financial Modelling*).
- [6] CORAZZA, M., MALLIARIS, A. G. e NARDELLI, C. (1996) Searching for Fractal Structure in Agricultural Futures Market, lecture presentata all'*Euro Working Group on Financial Modelling*, 14th Meeting, Mantova (sottoposto a *The J. of Futures Market*).
- [7] CORAZZA, M. e NARDELLI, C. (1993) Fenomeno della Dipendenza a Lungo Termine nel Mercato Finanziario Italiano, *Atti del XVII Convegno A.M.A.S.E.S.*, Ischia, 359-372.
- [8] FELDMAN, K. e KINGDON, J. (1995) Neural Networks and some Applications to Finance, *Applied Mathematical Finance*, 2, 17-42.
- [9] FIELITZ, B. D. e ROZELLE, J. P. (1972) Stable Distributions and the Mixtures of Distributions Hypotheses for Common Stock Returns, *J. of the American Statistical Association*, 78, 381, 28-36.
- [10] FIELITZ, B. D. e SMITH, E. W. (1972) Asymmetric Stable Distributions of Stock Price Changes, *J. of the American Statistical Association*, 67, 340, 813-814.
- [11] GATELY, E. (1996) *Neural Networks for Financial Forecasting*. John Wiley & Sons, New York.
- [12] GREENE, M. T. e FIELITZ, B. D. (1977) Long-term Dependence in Common Stock Returns, *J. of Fin. Econ.*, 4, 339-349.
- [13] HECHT-NIELSEN, R. (1990) *Neurocomputing*. Addison-Wesley, Reading.
- [14] HELMS, B. P., KAEN, F. R. e ROSENMAN, R. E. (1984) Memory in Commodity Futures Contracts, *The J. of Futures Markets*, 4, 559-567.
- [15] HINICH, M. J. e PATTERSON, D. M. (1989) Evidence of Nonlinearity in the Trade-by-trade Stock Market Return generating Process, in BARNETT, W.A., GEWEKE, J. e SHELL, K. (Eds.) (1989) *Economic*

- Complexity. Chaos, Sunspot, Bubbles and Nonlinearity.* Cambridge University Press, Cambridge.
- [16] HSIEH, D. A. (1991) Chaos and Nonlinear Dynamics: Application to Financial Markets, *The J. of Finance*, XLVI, 5, 1839-1877.
- [17] LO, A. W. (1991) Long-term Memory in Stock Market Prices, *Econometrica*, 59, 5, 1279-1313.
- [18] MALLIARIS, A. G. e PHILIPPATOS, G. (1992) Random Walk vs. Chaotic Dynamics in Financial Economics, lecture presentata al *Citicorp Seminar*, Brown.
- [19] MARGARITA, S. (1992) A Model of Non-learning Behavior in Artificial Economic Agents, *Atti del XVI Convegno A.M.A.S.E.S.*, Treviso.
- [20] MANDELBROT, B. B. (1963) The Variation of Certain Speculative Prices, *J. of Business*, in COOTNER, P. H. (Ed.) (1964) *The Random Character of Stock Market Prices*, 307-332, The M.I.T. Press, Cambridge, Massachusetts.
- [21] PETERS, E. E. (1991) *Chaos and Order in the Capital Market*. J. Wiley & Sons, New York.
- [22] PETERS, E. E. (1994) *Fractal Market Analysis*. J. Wiley & Sons, New York.
- [23] REFENES, A.-P. (Ed.) (1995) *Neural Networks in the Capital Markets*. J. Wiley & Sons, New York.
- [24] SANDULESCU, G., LAZAR, S. and CARDINALI, S. (1995) Neural Networks Applications in Finance. The Need of a Standardization of the Evaluation of the Performance related to Forecasting and Classification. Proposal and the Relation with Classical Methods, lecture presentata al *Financial Applications of Neural Networks and Fuzzy Systems*, 1st Meeting, Venice.
- [25] SCHEINKMAN, J. A. and LeBARON, B. (1989) Nonlinear Dynamics and Stock Returns, *The J. of Business*, 62, 3, 311-337.
- [26] SIMKOWITZ, M. and BEEDLES, W. (1980) Asymmetric Stable Distributed Security Returns, *J. of the American Statistical Association*, 75, 306-312.

Finito di stampare nel mese
di LUGLIO 1996
presso le Industrie Grafiche
ERREBI srl - Falconara/AN