Algoritmi Genetici per l'Evoluzione di Modelli Lineari
Metodologia ed Applicazioni

di

Marcello Galli*
Tommaso Minerva**

Novembre 1999

* Università degli Studi di Modena e Reggio Emilia
Dipartimento di Economia Politica
Via Berengario,51
41100 Modena (Italia)
e-mail:galli@mail.omnitel.it

** Università degli Studi di Modena e Reggio Emilia
Dipartimento di Economia Politica
Via Berengario,51
41100 Modena (Italia)
e-mail:minerva@unimo.it

Algoritmi Genetici per l' Evoluzione di Modelli Lineari
Metodologia ed Applicazioni

Marcello GALLI e Tommaso MINERVA

Riassunto

In questo lavoro vengono proposte alcune tecniche computazionali per il riconoscimento di modelli statistici cercando di fornire una risposta alla problematica della selezione del "miglior" modello statistico da utilizzare in una indagine predittiva. Vengono analizzate le tre fasi della selezione di un modello statistico: selezione delle variabili, individuazione della relazione funzionale tra le variabili e dei parametri a questa associati e stima del valore dei parametri. La selezione delle variabili procede, solitamente, mediante tecniche step-by-step in cui di volta in volta viene introdotta o eliminata una variabile alla volta valutandone, mediante criteri opportuni, la capacità esplicativa. Nel lavoro qui presentato viene proposto un approccio alternativo. Utilizzando tecniche numeriche proprie dell'Intelligenza Artificiale (algoritmi evolutivi, adattivi, decisionali) ed in particolare utilizzando algoritmi ibridi (integrazione tra Algoritmi Genetici e Logica Fuzzy) sono state affrontate le seguenti problematiche: come scegliere le variabili da utilizzare nell'ambito di un modello lineare multivariato, come selezionare i regressori nell'ambito di un modello AR ed ARMA ed infine auale criterio utilizzare per confrontare la bontà di due o più modelli. Per le prime due sono state proposte soluzioni basate sugli Algoritmi Genetici mentre per la terza si è proposto un criterio di valutazione basato sulla Logica Fuzzy. Le tecniche sono state valutate su insiemi di dati sperimentali o su dati simulati di cui si è fornito l'algoritmo di simulazione utilizzato. Un attento confronto con le tecniche standard è stato condotto nei casi più significativi evidenziando un significativo aumento delle performances predittive con l'utilizzo delle tecnologie implementate nell' ambito dello studio descritto da questa trattazione. Infine si è proceduto ad applicare gli algoritmi sviluppati a due casi reali: Previsioni del Livello di Marea della Laguna Veneziana e Analisi della Disoccupazione in Emilia Romagna evidenziando notevoli miglioramenti rispetto ad analoghe indagini condotte con tecniche classiche.

1. TECNICHE DI SELEZIONE DI UN MODELLO STATISTICO

La fase più operativa della statistica è diretta alla costruzione di un modello per descrivere, comprendere, prevedere, simulare e controllare un fenomeno reale. Per tali finalità diviene centrale la struttura logica e formale del modello di regressione mediante il quale si esplicita una relazione funzionale tra ciò che si intende spiegare (l'effetto, la risposta, il risultato) e quello che può esserne causa. Un modello statistico è una rappresentazione semplificata, analogica e necessaria della realtà derivata da osservazioni sperimentali oltre che da deduzioni logiche. L'aspetto dialettico nella costruzione di un modello statistico deriva dalle opposte esigenze di semplificare la struttura senza perdere in fedeltà, e tale conflitto è ineliminabile. Infatti, tutti i modelli sono intrinsecamente sbagliati: essi sono parzialmente e provvisoriamente utili, e sono destinati a essere sostituiti con l'avanzare del progresso scientifico e l'affinamento della conoscenza. Ciò che realmente conta non è la validità antologica delle relazioni accertate ma l'efficacia comparata in rapporto agli obiettivi. È l'obiettivo, infatti, che rende utile, efficace e temporaneamente valido il modello.

La costruzione di un modello si concretizza attraverso fasi successive: specificazione, stima e verifica del modello statistico. Non è un caso che la specificazione sia l'aspetto più delicato e importante dell'intera procedura; dalla sua correttezza, infatti, dipendono la validità e l'efficacia di tutte le fasi successive.

La specificazione di un modello statistico inizia con l'esplicitare un legame tra i fenomeni di interesse nel modo seguente:

$$(1) \qquad Y=f(X_1,X_2,\ldots X_p)$$

dove Y è la variabile da spiegare (variabile dipendente) mentre $X_1,X_2,\ldots X_p$ (variabili esplicative) sono variabili prescelte per spiegare Y, grazie alla funzione f.

Tale relazione deriva dalla interazione tra conoscenze a priori e risultati sperimentali, e poiché non esistono settori scientifici ove è lecito ipotizzare un legame di natura strettamente deterministica, a tale relazione dovrà essere associata una componente stocastica. Pertanto (nell'ipotesi di componente additiva dell'errore) la specificazione di un modello statistico spesso assume la forma:

$$(2) \qquad Y=f(X_1,X_2,\ldots X_p)+\varepsilon$$

in cui è necessario fare opportune ipotesi sulla distribuzione di ε.

Attente procedure di specificazione del modello sono essenziali per un corretto impiego delle informazioni empiriche disponibili. La ricerca di specificazione del modello si articola in diverse fasi: scelta della forma funzionale (lineare o non lineare e, in quest' ultimo caso, di quale classe di funzioni), scelta delle variabili (lista delle variabili d'interesse e distinzione fra variabili endogene e esogene) e individuazione del corretto numero di variabili.

Tale ricerca solleva importanti problemi circa la scelta delle variabili esplicative per un modello (problema dell'individuazione del corretto modello statistico). Diviene fondamentale, infatti, stabilire univocamente quali variabili sono significative e quali ridondanti per spiegare il fenomeno Y, all'interno del quadro teorico accettato dagli studiosi di quel determinato settore.

Caratteristica comune alle varie tecniche utilizzate in questo tipo di ricerca è la definizione aprioristica dei tre seguenti elementi:

1) definizione del modo per muoversi nello spazio dei modelli;
2) specificazione di un criterio di valutazione del modello;
3) modalità del termine della ricerca;

Negli anni si sono affermate diverse strategie (vedi [DrSm66]) per effettuare la scelta del modello "migliore". Tra queste citiamo:

a) Tutti i modelli possibili.
b) Backward elimination.
c) Forward selection.
d) Stepwise regression.
e) Stagewise regression.

La problematica inerente alla selezione del modello non si riferisce soltanto alla scelta delle dimensioni di un modello, ma anche a come valutare differenti modelli e sulla base di quali indicatori preferirne uno piuttosto che un altro.

La ricerca non può essere omologata a test di ipotesi, così come avviene nella fase di verifica del modello statistico, in quanto deriva da ipotesi separate, cioè da famiglie di distribuzioni di cui una non è un caso particolare dell'altra.

Si preferisce piuttosto, dati più modelli completamente generali per gli stessi dati, minimizzare un opportuno indice, espresso in funzione dei parametri del modello, tenuto conto di due comportamenti opposti: al crescere del numero dei parametri di un determinato modello la varianza dei residui diminuisce (poiché migliora l'adattamento), ma aumentano i vincoli imposti dagli stessi parametri (e quindi peggiora la parsimonia). I criteri di selezione cercano quindi di rappresentare una soluzione bilanciata tra complessità (maggior adattamento) e parsimonia (maggior rappresentatività delle variabili). In linea di massima si tratta di funzioni di verosimiglianza penalizzate, in cui all' aumentare della complessità del modello si associa un termine di penalizzazione crescente.

Rientrano a pieno titolo in questa filosofia i seguenti indicatori:

- ***L'indice di determinazione multipla corretto*** (***adjusted R^2***) introdotto nel 1961 da Theil [The61]:

$$\overline{R}^2 = 1 - \frac{\frac{ESS}{n-p-1}}{\frac{TSS}{n-1}} \quad ; \tag{3}$$

Dove:

$ESS = \sum_1^n (y_i - \hat{y}_i)^2$ (devianza dei residui = Error Sum of Squares);

$TSS = \sum_1^n (y_i - \bar{y})^2$ (devianza totale = Total Sum of Squares);

n = numero di osservazioni;

p = numero di variabili;

- ***L'indice di Mallows*** introdotto nel 1973 [Mal73] e definito da:

$$C_p(p) = \frac{RSS_j}{S_p^2} - (n - 2p) \; ; \tag{4}$$

dove:

RSS_j = somma dei quadrati dei residui (Regression Sum of Squares) o devianza della regressione;

p = numero dei parametri del modello;

n = numero di osservazioni;

$S_p^2 = \frac{\sum_1^n (\hat{E}_i)^2}{n-p-1}$ = varianza stimata sui residui in un modello con p parametri.

$\hat{E}_i = Y - XB$;

$B = (X'X)^{-1}X'Y$ (stimatore dei minimi quadrati);

- ***L'indice AIC*** (***Asymptotic Information Criterion***) proposto nel 1973 da Akaike [Akai73] e così definito:

(5) $$AIC(p) = n\log(S_p^2) + 2(p);$$

Nonostante di questo indicatore esistano varianti certamente preferibili e sia stato spesso criticato poiché sovraparametrizza il modello "ottimale", rimane sempre quello più utilizzato.

- ***L'indicatore AICC*** (***Corrected AIC***) presentato da Hurvich e Tsai nel 1989 [HuTs89] che rappresenta l'indice AIC corretto:

(6) $$AICC(p) = n\log(S_p^2) + 2\frac{n}{n-p-2}(p);$$

- ***L'indice BIC*** (***Bayesian Information Criterion***) proposto anche questo da Akaike nel 1978 [Akai78] definito da:

(7) $$BIC(p) = (n-p)\log\left(\frac{nS_p^2}{n-p}\right) + p\log\left(\frac{n(S_0^2 - S_p^2)}{p}\right);$$

che rappresenta una modifica dell'indice AIC per tener conto della riduzione di S_p^2 rispetto a S_0^2 (dove S_0^2 indica la varianza delle n osservazioni).

- ***L'indicatore di Schwarz*** [Schw78] proposto nel 1978 e approssimativamente equivalente al criterio BIC [Prie81]:

(8) $$SIC(p) = n\log(S_p^2) + p\log(n);$$

- ***L'indice RIS*** proposto nel 1978 da Rissanen [Rissa78] e così definito:

(9) $$RIS(p) = n\log(S^2) + (p+1)\log(n+2) + 2\log(p+1);$$

- ***L'indicatore di HAN*** proposto da Hannan e Quinn nel 1979 [HaQu79]:

(10) $$HAN(p) = n\log(S^2) + 2p\log(n);$$

Tutti i precedenti criteri sono asintoticamente equivalenti[1], con una tendenza alla sovraparametrizzazione per l'AIC e una tendenza alla sottoparametrizzazione per l'HAN.

Ritornando alle altre fasi per la costruzione del modello statistico bisogna osservare che la stima dei parametri di un modello può essere affrontata riconducendo tale problema allo schema più consolidato della teoria della stima mentre la verifica del modello statistico si concretizza in una serie articolata di decisioni inferenziali, spesso formalizzabili mediante il test delle ipotesi, orientate alla discussione critica del risultato ottenuto nella fase di stima. Se la verifica non conduce al rifiuto del modello stimato, allora tale modello può essere utilmente applicato; in caso contrario, occorre ripercorrere le tappe di specificazione-stima-verifica alla ricerca di un modello più soddisfacente.

2. TECNICHE COMPUTAZIONALI

L'enorme sviluppo dell'universo computazionale ha favorito l'affermarsi delle discipline appartenenti al campo del "Soft Computing" nel tentativo di creare sistemi automatici in grado di elaborare delle informazioni a supporto di processi decisionali, individuando quindi caso per caso la strategia da seguire. È allora in questo contesto che si può anche parlare di una sorta di "*Intelligenza Artificiale*"[2], evitando però di enfatizzare il termine *Intelligenza*. Non siamo infatti di fronte a una sorta di "miracolo" umano. Siamo semplicemente di fronte a un modo nuovo di "vedere" le cose, a una nuova possibile strada da seguire per giungere alla soluzione "ottima" di un dato problema, strada che sembra voler riproporre l'itinerario seguito dalla mente umana (anche se bisogna tenere ben presente che il cervello umano è una realtà ancora non del tutto conosciuta). I passi che si stanno muovendo portano allora verso tentativi di replicare il contesto decisionale umano (Logica *Fuzzy*) nel tentativo di considerare le diverse sfumature del mondo ed il contesto evolutivo umano (*Algoritmi Genetici*) dove si afferma il concetto di sopravvivenza dell'individuo migliore, introducendo, quindi, il concetto di elaborazione parallela e non più seriale. Queste tecnologie sono state e vengono anche oggi utilizzate in ambito statistico in contrasto con le più convenzionali tecniche classiche.

Gli Algoritmi Genetici [Holl75, Gold89, Davi91, Mit96] costituiscono una classe di tecnologie di ricerca probabilistica che si ispirano all'evoluzione biologica. L'algoritmo fornisce uno strumento per l'interrogazione di insiemi molto estesi di dati e per analizzare relazioni funzionali complesse. Il processo di ricerca si basa su una versione simulata dell'evoluzione in senso darwiniano, nella quale una popolazione di soluzioni candidate vengono manipolate e condizionate, per mezzo di una strategia artificialmente generata di sopravvivenza dei soggetti più adatti dal punto di vista evolutivo.

Gran parte del lavoro effettuato dagli statistici viene dedicato alla costruzione di modelli, alla stima dei loro parametri, nonché alla convalidazione dei modelli stessi. La maggior parte del lavoro della statistica applicata viene intrapreso attraverso considerazioni e restrizioni di tipo matematico e computazionale . Così, per esempio, continuità e differenziabilità della forma funzionale di un modello sono requisiti artificiali imposti per una più facile e pronta disponibilità dei metodi di stima. Gli Algoritmi Genetici rimuovono tali ipotesi permettendo di trovare soluzioni ottimali in quasi totale assenza di restrizioni imposte. Qui studieremo in particolare il loro comportamento nei problemi di identificazione di modelli di regressione lineare e di modelli autoregressivi lineari.

---

[1] Anche se sono parecchi gli statistici che ritengono più validi i criteri SIC e BIC.

[2] Con il termine "Intelligenza Artificiale " si intende di norma una vera e propria disciplina atta a studiare i fondamenti teorici, i metodi principali, i criteri di progettazione e la costruzione di programmi che possano permettere al calcolatore di svolgere attività tipicamente umane, e di renderlo capace di sostituirsi allo stesso uomo nell'affrontare una molteplicità di problemi, nello stesso modo in cui egli li avrebbe affrontati, ottimizzando però il raggiungimento dei risultati.

Nei Sistemi Fuzzy [Zade84, DuPr80] viene abbandonata la tipica logica matematica bivalente, dove non esistono vie di mezzo tra il vero e il falso, il bianco e il nero, il giusto e lo sbagliato, a favore di una logica multivalente in cui sono ammesse le varie sfumature tra i due estremi.

Tale logica, che discende dalla Teoria degli Insiemi Sfumati, fornisce un meccanismo per consentire ai sistemi preposti ai processi decisionali di gestire informazioni caratterizzate da un basso grado di precisione. Se ne desume che tale logica non può essere un algoritmo né una classe di algoritmi come nel caso delle due precedenti tecnologie. Come tale, non costituisce in modo intrinseco uno strumento di automazione, ma piuttosto serve a facilitare la definizione degli ambiti di determinate forme di processo decisionale che potrebbero rendersi necessarie quando un sistema si trova a dover interagire con il mondo esterno. In particolare, la Logica Fuzzy costituisce uno strumento flessibile attraverso il quale un sistema può allo stesso tempo ricevere istruzioni e dare spiegazioni all'utente sulle azioni intraprese. Un loro approccio verrà considerato soltanto allo scopo di definire, per un determinato modello statistico, quali e quanti parametri considerare sulla base di varie e contrastanti informazioni e valutazioni.

3. SELEZIONE DI UN MODELLO LINEARE

In questa sezione viene presentato un esempio di modello lineare cercando di affrontare la problematica inerente alla sua specificazione sia attraverso le tecniche classiche più comuni, sia attraverso tecniche computazionali "intelligenti".

Il problema può essere posto nei seguenti termini:

dato l'insieme $X \equiv \{x_1, x_2, ..., x_n\}$ di variabili indipendenti, di cui disponiamo k osservazioni, e la variabile dipendente Y, si tratta di determinare tra tutti i possibili sottoinsiemi propri e impropri di X (che sono , escludendo il solo modello vuoto, $2^n - 1$) quello che meglio degli altri specifica il modello in esame.

Se $Y = f(\tilde{x}, \Theta, \varepsilon)$ rappresenta il modello scelto, che in ipotesi di additività dell'errore ε può scriversi $Y = f(\tilde{x}, \Theta) + \varepsilon$, in cui $\tilde{x}$ rappresenta l'insieme delle variabili indipendenti scelte per "spiegare" la Y. Supponendo ora che la relazione f che lega le variabili indipendenti a quella dipendente sia tipo lineare (univariato o multivariato) il problema successivo sarà quello di determinare i coefficienti dei parametri Θ attraverso le solite procedure dei minimi quadrati o di massima verosimiglianza.

Quindi, se in generale il problema prevede la risoluzione dei tre seguenti sottoproblemi:

- selezionare f
- selezionare $\tilde{x}$
- determinare Θ

nel caso specifico sarà determinante la selezione di $\tilde{x}$ (date le ipotesi fatte su f, sul calcolo di Θ e sulla distribuzione della componente stocastica).

A questo proposito si è osservato che le difficoltà da risolvere riguardano sia il numero di variabili da scegliere (o meglio quali variabili scegliere) sia i criteri coi quali effettuare questa scelta. Basiamo le nostre considerazioni su un insieme di dati relativo a misure di spettri di assorbimento a 21 frequenze diverse di radiazione elettromagnetica di 264 campioni ematici. A queste misure viene associata la misura dei diversi indici del livello di colesterolo nel sangue. In particolare supponendo di voler tarare uno strumento che da misure di assorbimento (variabili indipendenti) riesca a determinare il livello di colesterolo (variabile dipendente, Y) ci si chiederà

"Quali sono le frequenze tra le 21 disponibili che determinano il valore di Y ?". Il dataset completo e' disponibile all' interno della Toolbox Statistics di Matlab. Per poter rispondere a tale domanda è necessario spiegare le tecniche attraverso le quali si è cerca di risolvere il problema.

Come primo approccio è stato utilizzata una tecnica "Stepwise" con controllo interattivo [Bra97] che permette in modo manuale di selezionare o eliminare dall'insieme complessivo di variabili a disposizione quelle più o meno significative valutate sulla base del *p-value* valutato con un t-test.

Purtroppo questo tipo di procedura, come detto, funziona manualmente e non permette assolutamente di generare in modo completamente automatico l'insieme di variabili rilevanti (cioè, nel caso specifico, quelle frequenze che sono in grado di determinare il valore del livello di colesterolo nel sangue). Successivamente sono state valutate e confrontate tecniche stepwise automatizzate. E' necessario premettere, pero', che questo campione di dati presenta problemi di multicollinearità (cioè le variabili esplicative sono fra loro fortemente correlate) rendendo il confronto fra le diverse tecniche ancora più interessante.

Backward elimination method: questa procedura ad ogni iterazione scarta tra i regressori non significativi quello meno significativo, fermandosi quando tutte le variabili rimaste sono significative (vedi [DrSm66]). Con la stessa tecnica di procedimento (dal modello più ampio al modello più ristretto) si è poi cercato di generalizzare tale metodo al fine di poter utilizzare altri criteri. Si è così sviluppato un algoritmo che prevedesse l'eliminazione di quelle variabili che, una volta "tolte", migliorano più delle altre il modello in base a un determinato indicatore.

Gli indicatori utilizzati sono quelli esposti nelle sezioni: R^2, $\overline{R}^2$, *AIC, AICC, BIC, SIC, RIS e HAN.* Criterio di stop sarà quindi il peggioramento del modello (sulla base di tali indici) all'ulteriore "mossa" successiva.

Forward selection method: questa tecnica consiste nell'inserire una variabile per volta al modello a seconda del suo coefficiente parziale di correlazione fino a che la successiva variabile inserita risulti essere non significativa (vedi [DrSm66]). Allo stesso modo di quanto fatto per il programma "Backward", anche in questo caso si è cercato di utilizzare questa procedura anche attraverso altri criteri. L' approccio in questione prevede di inserire in modo iterativo la variabile che più di ogni altra migliora la "bontà" del modello in base all'indicatore prescelto. Il processo avrà termine quando, inserendo qualunque ulteriore variabile, il modello ha comunque un peggioramento complessivo.

L'intero "data-set" (264 campioni di sangue) è stato suddiviso in tre parti:

- Training set (l'esatta metà dei campioni di sangue);
- Validation set (un quarto dell'intero set di dati);
- Test set (la rimanente parte).

Nonostante diverse siano state le modalità attraverso cui sono state suddivise le percentuali di questi campioni (provando a "rimescolare" l'intero data-set ma sempre attribuendo il 50% dei dati al training set e il 25% rispettivamente al validation e al test set) non si sono avvertiti significativi cambiamenti nei risultati mostrati.

Per quanto riguarda il "training" set si può dire che esso costituisce quell'insieme di dati sul quale si è effettuata la regressione lineare con lo scopo di calcolare i coefficienti delle variabili selezionate.

Il "validation" invece costituisce quell'insieme di dati sul quale si sono calcolati gli indici utilizzati come criterio di valutazione della bontà del modello.

Sul "test" set invece si è valutato l' R^2 della regressione e gli indici di adattamento.

I risultati ottenuti sono indicati nelle due tabelle di seguito riportate in cui nella prima colonna della tabella, "metodo" indica il tipo di programma utilizzato per ottenere i risultati, nella seconda colonna "criterio" indica quale indice di riferimento si è utilizzato (*p-value* indica semplicemente

che si è utilizzato il metodo "backward elimination" su esposto, *P_corr* invece indica l'utilizzo del "Forward selection" puro), nella terza colonna è stato inserito l' R^2 calcolato sul test set, nella quarta colonna viene riportato il valore della stima dello scarto quadratico medio previsivo, nella quinta colonna è evidenziato il numero di variabili scelte dal programma per quel tipo di criterio, *NV*, (dimensione del modello ottimo) e nell'ultima colonna viene indicato quante variabili tra quelle scelte risultano statisticamente non significative sulla base del t-test, NV_t.

Per quanto concerne i criteri: *p-value* e *P_corr* sono già stati spiegati, *Theil* è l'indice di determinazione multipla corretto o $\overline{R}^2$, *RSQ* rappresenta l' R^2, *AIC* è l'indice di Akaike (Asymptotic Information Criterion), *AICC* è l'indicatore *AIC* corretto, *BIC* coincide con il "Bayesian Information Criterion", *SIC* sta a indicare l'indice di Schwarz, *RIS* quello di Rissanen e *HAN* quello proposto da Hannan e Quinn.

METODO	CRITERIO	RSQ_{test}	MSE_p	NV	NV_t
Backward	*p-value*	0.788	5.85	12	-
Backward	*Theil*	0.808	5.36	14	5
Backward	*RSQ*	0.808	-	18	11
Backward	*AIC*	0.781	4.62	11	6
Backward	*AICC*	0.781	4.62	11	6
Backward	*BIC*	0.781	4.62	11	6
Backward	*SIC*	0.781	4.62	11	6
Backward	*RIS*	0.781	4.62	11	6
Backward	*HAN*	0.746	4.82	7	-
Forward	*P_corr*	0.736	4.76	2	-
Forward	*Theil*	0.757	4.24	2	-
Forward	*RSQ*	0.763	4.22	3	1
Forward	*AIC*	0.757	4.24	2	-
Forward	*AICC*	0.757	4.24	2	-
Forward	*BIC*	0.757	4.24	2	-
Forward	*SIC*	0.757	4.24	2	-
Forward	*RIS*	0.757	4.24	2	-
Forward	*HAN*	0.757	4.24	2	-

Tabella 1. Risultati ottenuti con la tecnica di selezione Stepwise sia Backward che Forward sul campione di dati considerato.

Nella tabella 2 sono invece rappresentati i risultati ottenuti con una procedura di "Stepwise" interattivo e i dati relativi al modello saturo (completo di tutte le 21 frequenze).

METODO	CRITERI	RSQ_{test}	MSEP	N.V.	N_t
Stepwise	MANUALE	0.736	4.76	2	-
Modello completo	----------	0.790	5.94	21	14

Tabella 2. Risultati ottenuti con la tecnica "Stepwise" (eliminazione manuale delle variabili) e con il modello completo.

Dall'esame di questi risultati si possono trarre alcune conclusioni: la presenza di variabili fortemente correlate rende sicuramente preferibile in generale i risultati ottenuti con la tecnica "forward" rispetto a quelli ottenuti con la tecnica "backward". Infatti, nonostante l' R^2 presenti valori leggermente più bassi, la dimensione del modello è decisamente inferiore (e questo in parte spiega un minore R^2) con la totalità delle variabili statisticamente significative (eccetto il caso in cui si è scelto come criterio di ottimizzazione l' R^2 , che presenta una sola variabile non

significativa in base al test statistico t a 0.95). Inoltre anche lo scarto quadratico medio è decisamente minore a conferma della maggiore "bontà" dei modelli.

Premesso questo, analizziamo i risultati dei diversi criteri di scelta:

- Con le tecniche "backward" la scelta di utilizzare il p-value come criterio, e quindi di giudicare la bontà del modello sulla base della sola significatività delle variabili in base al test-t, non risulta convincente in quanto lo scarto quadratico medio previsivo della regressione risulta assai più alto rispetto agli altri (MSE = 5.85) e il numero di variabili selezionate non è inferiore (N.V. = 12) sebbene queste siano ovviamente tutte significative. Il criterio di scelta complessivamente migliore per l'utilizzo di questo programma sembra essere, nonostante la presenza di ben sei variabili non significative sulla base del test-t, il metodo che sfrutta come criterio di scelta l'indice *HAN* che seleziona un modello di dimensione inferiore e con uno scarto quadratico minore. Per quanto riguarda gli indicatori *AIC, AICC, BIC, SIC e RIS* i risultati sono i medesimi.

- Utilizzando la tecnica "forward" , come già detto, i risultati migliorano.

Confrontando i singoli criteri bisogna osservare che nella maggior parte di essi i risultati sono tutti uguali con eccezione del caso identificato come "*P_corr*" che risulta essere il peggiore sia in termini di R^2 che di scarto quadratico medio previsivo. L' R^2 come criterio di scelta permette di ottenere risultati leggermente migliori sia in termini di R^2 che di scarto quadratico medio previsivo.

Concludendo possiamo affermare che, utilizzando le tecniche classiche modificate per introdurre la scelta attraverso indicatori differenti da quelli generalmente proposti nelle tecniche pure di "backward elimination" e "forward selection" ("*p-value*" nel "backward" e "*P_corr*" nel "forward"), si ottengono risultati migliori. Inoltre, i risultati evidenziati nella tabella 2 mostrano che le procedure automatizzate qui implementate migliorano i risultati sia rispetto al modello completo (prevedibile a causa della forte correlazione tra le variabili) sia rispetto a una procedura "Stepwise" interattiva.

3. 1. IL MODELLO LINEARE GENETICAMENTE EVOLUTO

Tutte quante le tecniche classiche finora presentate soffrono di due difetti non trascurabili che le rendono fortemente criticabili:

1. Includono (o escludono) tutte, infatti, una variabile per volta "costruendo" un modello finale in funzione dell'ordine della sequenza attraverso la quale viene fatta la scelta.Se si confrontano i risultati ottenuti con le due tecniche viste è possibile rendersi immediatamente conto di quanto questa "dipendenza dalla sequenza" sia importante. Multicollinearità a parte, infatti, a priori sarebbe stato lecito aspettarsi i medesimi risultati con lo stesso criterio per entrambe le procedure. Ma così non è stato.
2. Viene meno la possibilità da parte dell'operatore (avendo automatizzato la procedura) di selezionare manualmente quelle variabili che, indipendentemente dal peggioramento o miglioramento globale del modello, la teoria sottostante giudichi imprescindibili.

La necessità di costruire un algoritmo efficiente e che permetta di selezionare un modello che non dipenda dalla sequenza delle variabili scelte, ci porta a provare a costruire un algoritmo genetico.

Gli algoritmi genetici [Holl75, Gold89, Davi91, Mit96] costituiscono una classe di tecnologie di ricerca probabilistica che si ispirano all'evoluzione biologica. La Fig. 1 rappresenta il ciclo evolutivo dell'algoritmo genetico che costituisce il meccanismo centrale di ricerca. L'algoritmo imita l'evoluzione naturale attraverso il cambiamento (o evoluzione) ripetuto di una popolazione di soluzioni candidate nel tentativo di trovare la soluzione ottimale.

L'evoluzione artificiale è paragonabile a un allevamento. Ci sono mutamenti casuali nella popolazione ma è l'allevatore a decidere di volta in volta quali sono le caratteristiche interessanti che devono essere selezionate. Questo tipo di evoluzione si dice pertanto diretta per distinguerla da quella naturale (indiretta) in cui gli individui sono migliori se in generale sopravvivono e si adattano ai mutamenti che possono avvenire nell'ambiente.

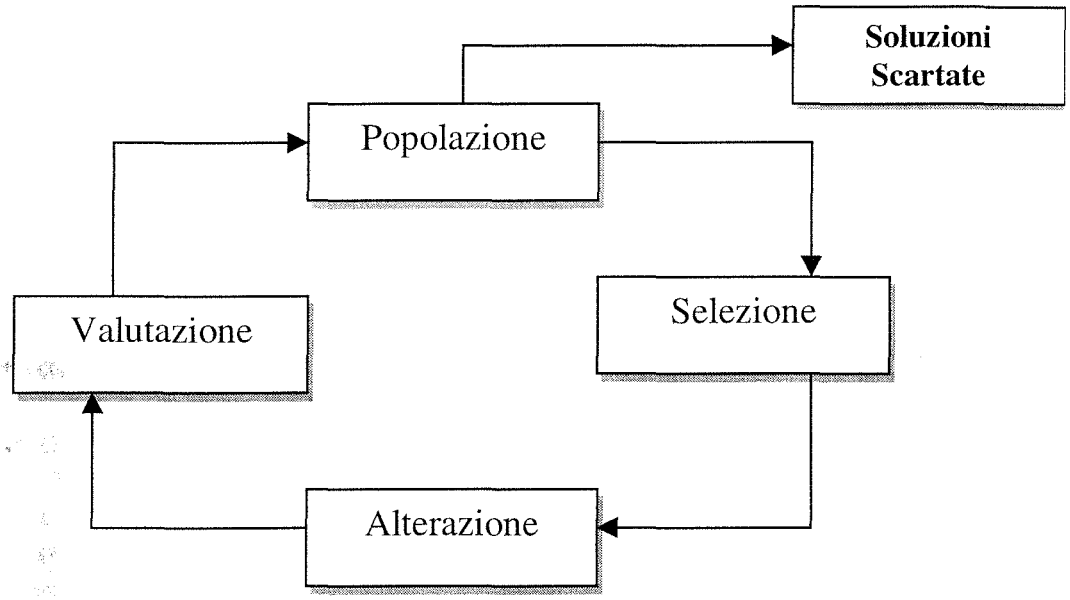

Figura 1. Ciclo evolutivo di un Algoritmo Genetico.

Scopo del nostro algoritmo è quello di determinare tra le 21 frequenze dei nostri campioni quali riescano a determinare il livello di colesterolo nel sangue.

Si codificheranno pertanto le 21 frequenze in una sola stringa, che rappresenta un potenziale individuo (cromosoma) della popolazione corrispondente a una determinata generazione dell'algoritmo, la quale, sulla base del principio darwiniano di sopravvivenza dell'individuo migliore, e quindi di quello che ottiene la migliore prestazione in base all'indicatore prescelto, evolverà e fornirà le variabili che meglio valutino il livello di colesterolo nel sangue.

Di conseguenza nella nostra rappresentazione a ogni individuo sarà associato un possibile modello statistico.

La rappresentazione dell'individuo è riportata in Fig. 2.

1	0	1	1	1	0	1	0	0	0	0	0	1	0	0	1	0	0	0	1	1

Figura 2. Rappresentazione dell'individuo utilizzato nel nostro algoritmo.

Ogni stringa è suddivisa in 21 celle (geni), ognuna delle quali può trovarsi in uno dei due stati acceso-spento: 0 o 1 (allele).

Ogni gene corrisponde a una frequenza diversa, se il corrispondente allele avrà valore 1 allora significherà che quella frequenza sarà inclusa nel nostro modello, se invece avrà valore 0 significherà che sarà esclusa.

All'individuo rappresentato in Fig. 2 corrisponderà, dunque, il seguente modello:

$$(11) \qquad Y = \beta_1 X_1 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_7 X_7 + \beta_{13} X_{13} + \beta_{16} X_{16} + \beta_{20} X_{20} + \beta_{21} X_{21}$$

La popolazione iniziale viene generata in modo casuale. Il numero di individui e di generazioni massime potrà essere definito a piacimento (nel caso in esame sotto riportato si è scelto di generare una popolazione di 100 individui con un numero massimo di generazioni pari a 100). Quando la popolazione è stata codificata può iniziare il ciclo evolutivo dell'algoritmo genetico.

Ogni stringa della popolazione viene poi valutata secondo una "funzione di Fitness" che indica il suo grado di adattamento e che costituisce il criterio in base al quale determinare la probabilità di sopravvivenza dell'individuo nella generazione successiva. La ricerca dell'individuo migliore sarà fatta cercando il cromosoma (stringa) con fitness maggiore.

Nel nostro caso, per ogni individuo della popolazione si procederà a stimare i parametri effettuando una regressione lineare sul campione di dati inserito nel training-set. Il passo successivo corrisponde al calcolo degli indici sul validation-set. Il valore di questi indicatori, normalizzati sull'intera popolazione in un intervallo compreso tra 0 e 2, costituisce il valore attribuito dalla nostra funzione di fitness. Una volta determinati i valori di fitness dei singoli individui, si procederà a manipolare geneticamente questi cromosomi.

È questa la cosiddetta fase della selezione: si costruisce una popolazione intermedia che, una volta applicate le procedure di clonazione, crossover (ricombinazione) e mutazione, fornirà la popolazione della generazione successiva. Gli individui cui corrisponde un valore di fitness migliore dovrebbero essere selezionati con una probabilità maggiore, nel pieno rispetto della teoria della selezione naturale. Il procedimento di selezione risulta di fondamentale importanza poiché è lo strumento attraverso cui si cerca di convergere verso la soluzione migliore.

Con la procedura di clonazione una frazione della popolazione, selezionata casualmente in base alle probabilità attribuite dai rispettivi valori di fitness, viene passata alla generazione successiva senza subire alcun cambiamento.

Il crossover consiste in una versione artificiale della riproduzione sessuata. Se due individui possiedono valori elevati di fitness, allora l'algoritmo esplora l'eventualità che una combinazione dei loro geni possa far scaturire una progenie con caratteristiche di fitness ancora maggiore. Gli individui con elevati valori di fitness hanno un'alta probabilità di riprodursi, mentre gli individui caratterizzati da fitness scarsa dovrebbero avere una bassa probabilità di riprodursi.

Se il crossover rappresenta il mezzo attraverso il quale spostarsi all'interno dello spazio in base alle informazioni passate, la mutazione rappresenta l'innovazione. In concreto si tratta di modificazioni casuali nella struttura genetica degli individui (ovviamente con un basso tasso di probabilità). La mutazione permette perciò di esplorare nuove aree della superficie di risposta.

Una volta applicati gli operatori biologici e generata una nuova popolazione, il processo viene ripetuto finché non converge (tutte le unità della popolazione saranno uguali) oppure finché non venga violato qualche parametro fisso di controllo (come per esempio un numero prefissato di generazioni o non venga trovato quell'individuo in grado di soddisfare l'obiettivo inizialmente prefissato).

Nel caso specifico in esame il criterio di ottimizzazione sarà soddisfatto soltanto quando si raggiungerà un numero di generazioni pari a 100 (che garantisce la convergenza così come verificato manualmente).

Gli altri parametri fissati nel programma sono le percentuali di mappatura dell'algoritmo. L'80% degli individui della popolazione (i migliori) vengono predisposti all'incrocio con altri individui. La probabilità di effettuare l'incrocio è ancora dell'80%. L'incrocio scelto è del tipo "single point

crossover" che porta all'effettuazione di un solo taglio per ogni stringa nelle ricombinazioni. Non ci sono individui clonati tranne quelli per cui non avviene l'incrocio, che passano alla generazione successiva. Il restante 20% di individui della generazione successiva viene generato e reinserito casualmente. La probabilità di una possibile mutazione nei cromosomi di ogni individuo è di circa il 3%.

Bisogna infine ricordare che per ottenere uno snellimento generale della procedura un semplificazione dei calcoli i dati in input (il dataset "choles") sono stati normalizzati nell'intervallo compreso tra –1 e 1.

Nella tabella 3 sono riportati i risultati ottenuti con questo programma[3].

METODO	CRITERIO	RSQ_{test}	MSEP	N.V.	N_t
GALMS	*RSQ*	0.804	5.13	15	7
GALMS	*AIC*	0.808	3.87	5	-

Tabella 3. Risultati ottenuti con il programma GALMS (Genetic Algoritm for Linear Model Selection).

Così come era prevedibile aspettarsi, otteniamo risultati decisamente migliori rispetto ai due programmi precedenti.

I problemi connessi alla forte correlazione tra le variabili esplicative, qui non influenzano in modo decisivo la dimensione del modello selezionato. Gli algoritmi genetici, lavorando in modo parallelo, ottimizzano una procedura che seleziona le variabili contemporaneamente e non in sequenza.

Il programma, inoltre, ottimizza il valore dell'indicatore scelto senza preoccuparsi del numero di variabili selezionate e della loro significatività. È questo il motivo per cui ci troviamo di fronte a modelli di dimensioni maggiori rispetto a quelli ottenuti con la tecnica "Forward" e di dimensioni inferiori rispetto a quelli ottenuti con la tecnica "Backward".

Comunque, l'aver implementato una procedura la cui ottimizzazione risulta essere indipendente dalla sequenza di scelta, permette di ottenere modelli più efficienti da un punto di vista predittivo.

Non possono certamente essere definiti sbalorditivi i risultati ottenuti con il criterio *RSQ* (soprattutto se confrontati con quelli ottenuti con il medesimo criterio utilizzando l' approccio "Forward"), ma certamente ottimi sono i risultati ottenuti con il criterio *AIC*.

Si ottiene, infatti, un modello con 5 variabili esplicative tutte significative con uno scarto quadratico previsivo medio inferiore di oltre mezzo punto rispetto a tutti i modelli esaminati finora.

Tale impostazione non elimina né affronta i problemi connessi all'automazione della procedura.

Abbiamo, infatti, sempre a che fare con un programma che lavora completamente in modo automatico senza tenere in alcuna considerazione la teoria sottostante che l'utente potrebbe voler imporre o vincolare.

Ricordiamo infatti che in generale il problema della selezione del miglior modello statistico richiede pur sempre un compromesso tra l'automazione di una procedura che sia universalmente utilizzabile e il giudizio personale e soggettivo dell'utente interessato a tale analisi.

Un tentativo di risolvere questo tipo di problematica è stato fatto utilizzando un algoritmo che deriva le condizioni di ottimizzazione sulla base delle conoscenze soggettive e oggettive possedute a priori dall'utente.

Per far ciò si è impiegata la cosiddetta "Logica Fuzzy" e si è costruito un criterio per la selezione di un modello che denomineremo "sfumato".

---

[3] Siccome questo tipo di procedura impiega parecchie risorse in termini di tempo, non si è sperimentata l'analisi per tutti gli indicatori utilizzati in precedenza, ma soltanto per i più rappresentativi. L'AIC si è rivelato un indicatore molto adatto (in quanto ha sempre ottenuto un MSEP più basso) a questo tipo di analisi.

3. 2. IL MODELLO GENETICAMENTE EVOLUTO "SFUMATO"

La Logica Sfumata (Fuzzy Logic) [Zade84, DuPr80] discende dalla "Teoria degli Insiemi Sfumati" [Zade65], intesa a sistematizzare la precedente idea di logica a valori infiniti, e fornisce un meccanismo per consentire ai sistemi preposti ai processi decisionali di gestire informazioni caratterizzate da un basso grado di precisione.

L'uomo è in grado di manipolare concetti imprecisi, come "abbastanza piccolo", "piuttosto grande", o "molto recente"; in concreto, la logica sfumata è un tentativo di consentire agli elaboratori elettronici di ragionare e contemporaneamente giustificare le proprie azioni in modo altrettanto flessibile. La logica tradizionale "Booleana" preclude in modo assoluto la appartenenza non totale a un insieme. La logica sfumata, invece, consentendo un'appartenenza parziale e specificando operazioni sugli insiemi basate sugli insiemi stessi così definiti, fornisce una struttura grazie alla quale operare una forma di ragionamento automatizzato molto più trasparente, e potenzialmente più raffinato.

La logica convenzionale utilizza insiemi con confini rigidi: questo significa che vi è una transizione istantanea da un insieme all'altro. Qualsiasi processo decisionale basato su questo tipo di confini deve pertanto gestire cambiamenti di stato bruschi e istantanei, che potrebbero divenire oltremodo artificiosi nella costruzione di modelli per il funzionamento del mondo reale.

La logica sfumata rimuove i confini rigidi attraverso l'uso di funzioni di appartenenza agli insiemi. Le funzioni di appartenenza indicano il grado di appartenenza a un insieme sfumato restituendo un valore che può variare tra 0 (per indicare un'appartenenza nulla) e 1 (per indicare appartenenza totale).

La Logica Sfumata, non può essere considerata, quindi, uno strumento di automazione né un algoritmo, ma piuttosto un mezzo attraverso cui poter facilitare la definizione degli ambiti di determinate forme di processo decisionale.

In tale contesto, si è pensato di costruire un indicatore (che si chiamerà *FIS*: "Fuzzy Inference System") ottenuto mediante l'utilizzo di questi criteri.

Esaminando i risultati ottenuti coi programmi "Backward", "Forward" e "GALMS", ci si può facilmente rendere conto della necessità di voler trovare un modello che presenti contemporaneamente le seguenti caratteristiche:

1) $R^2 \Longrightarrow$ alto (tendente a 1).
2) Numero di variabili del modello (dimensione) $\Longrightarrow$ basso .
3) Variabili non significative (con t-test < t_α) $\Longrightarrow$ basso (tendente a 0).

Lavorando nell' ambito dell' approccio Fuzzy si riesce facilmente a utilizzare come funzione di fitness un indicatore *(FIS)* che permetta di ottimizzare i tre parametri posti sopra (in antitesi tra loro) e di considerare tutte le altre valutazioni soggettive eventualmente poste dall'operatore.

Per comprendere meglio l'importanza di tutto questo si supponga di affrontare il problema della selezione delle variabili esplicative che "spieghino" il livello di altezza degli individui.

Quale giudizio di bontà si dovrebbe attribuire a un modello qualora, tra le variabili selezionate, non sia presente la variabile "altezza dei genitori" ?

È ovvio che avendo la possibilità di giudicare il modello scelto mediante un indicatore che ritenga vincolante ai fini del giudizio di bontà l'inserimento di questa variabile, anche procedure automatizzate non produrrebbero gli inconvenienti generalmente a queste associabili.

Questo è il motivo per cui si ritiene doveroso implementare questo tipo di procedura.

La Logica Fuzzy si basa sulla descrizione del grado di appartenenza delle variabili a un insieme Fuzzy. Pertanto, se ritorniamo all'esempio simulato del campione di dati del file "choles", la logica sfumata si occuperà di assegnare a ogni variabile di input, lungo l'intero spazio dei valori assunti dalla variabile stessa (il cosiddetto "Universe of Discourse"), vari insiemi sfumati.

Quindi, se l'Universe of Discourse della variabile di input R^2 è l'insieme [0,1], si procederà a suddividerlo in alcuni insiemi sfumati i cui confini precisi di transizione vengono rimossi (vedi a riguardo la Fig. 3). Agli insiemi sfumati così definiti vengono collegati quantificatori linguistici (nel caso della Fig. 3 : "BASSO", "MEDIO-BASSO", "MEDIO", "MEDIO-ALTO", "ALTO") allo scopo di rendere più agevole l'interpretazione dei risultati, oltre che per facilitare le modalità di definizione degli insiemi.Una volta definiti gli insiemi sfumati, qualsiasi valore discreto di R^2 misurato nel mondo reale può essere convertito in termini sfumati.

La Fig. 3 mostra come il valore 0.15 si posiziona sull'universo definito per la variabile "R^2". Utilizzando le funzioni di appartenenza, si ottengono appartenenza parziali per gli insiemi "BASSO" e "MEDIO-BASSO" con valori rispettivamente di 0.4 e 0.6 ($R^2(0.15)$ = 0.4 / BASSO , 0.6 / MEDIO-BASSO).

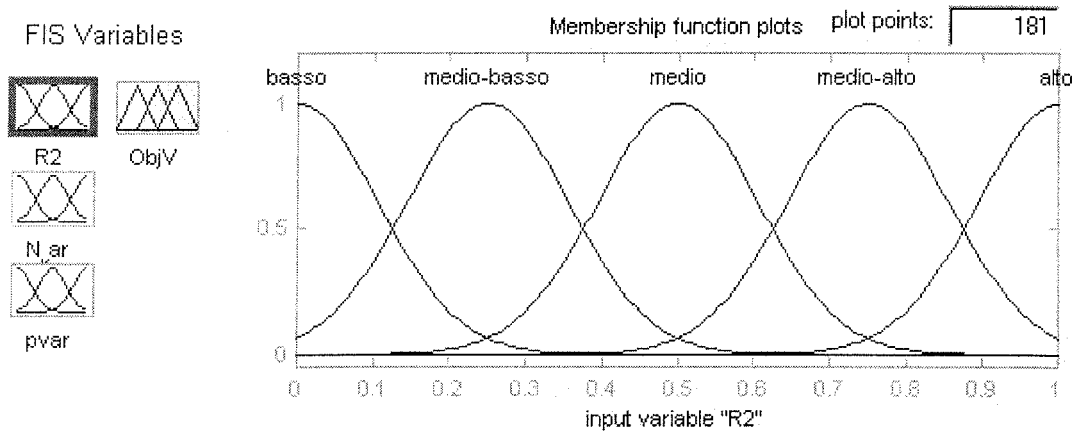

Figura 3. Universe of Discourse di R^2 e relativi insiemi sfumati, coi rispettivi quantificatori linguistici.

Lo stesso procedimento deve essere applicato alle altre variabili di input (numero di variabili del modello e variabili non significative) e alla variabile di output (l'indicatore di bontà del modello che restituisce un valore compreso tra 0 e 1).

Il processo di inferenza è poi definito tramite l'uso di una base di regole di produzione (relazioni logiche) nella forma "SE X E/O Y E/O ... ALLORA Z". Questa metodologia è direttamente compatibile con le basi di regole comunemente utilizzate per i sistemi esperti (con la sola differenza che nei sistemi esperti gli input devono ricadere sempre in un insieme definito e distinto che possa rendere vera la regola mentre in quelli sfumati possono appartenere a più di un insieme, rendendo vere determinate regole, ma "in una certa misura"). Per esempio: " Se R^2 è alto" E "numero di variabili è basso" E "numero di variabili non significative è nullo" ALLORA "il modello è ottimo". Dopo aver definito l'insieme complessivo delle 75 regole (5 x 5 x 3 : prodotto del numero di quantificatori linguistici utilizzati per le variabili di input) MATLAB restituisce una struttura (Fis.fis) che ricevendo in input il valore discreto di ciascuna delle tre variabili di input, restituisce il valore (come detto compreso tra o e 1) di confronto della bontà del modello.

Tale valore costituisce il criterio di scelta delle variabili esplicative del modello e verrà di seguito chiamato *FIS* (Fuzzy Inference System).

Il processo inferenziale tipico dei sistemi sfumati, quindi, consentendo di effettuare ragionamenti sulla base di tali appartenenze parziali e integrato da un modello di "desfumatura", tende a far ritornare il sistema a uno stato di valori discreti.

Applicando tale sistema ai tre programmi visti nei paragrafi precedenti si sono ottenuti i risultati mostrati nella tabella 4.

METODO	CRITERIO	RSQ_{test}	MSEP	N.V.	N_t
Backward	*FIS*	0.799	5.27	13	4
Forward	*FIS*	0.757	4.24	2	-
GALMS	*FIS*	0.772	4.12	2	-

Tabella 4. Risultati ottenuti con il criterio *FIS* sul campione di dati utilizzato.

Molto interessante è l'analisi dell'andamento dell'indice *FIS* al variare degli indicatori utilizzati per costruirlo (per poter rappresentare tale andamento su un grafico a tre dimensioni si è dovuto porre il numero di variabili complessive del modello costante, vedi Fig. 4).

Il ciclo iterativo dei programmi ha come criterio di stop l'eventuale peggioramento del modello e, se si parte dal modello nullo (programma "Forward"), la forte correlazione presente tra le variabili esplicative tende a farsi sentire immediatamente (essendo l'indice stesso costruito su tali informazioni) peggiorando il modello stesso. In questo modo il valore dell' R^2 non fa in tempo a crescere che l'algoritmo ha terminato il ciclo. Questo è il motivo per cui l' R^2_{test} (RSQ_{test}) è relativamente basso nel programma in questione rispetto al programma "Backward".

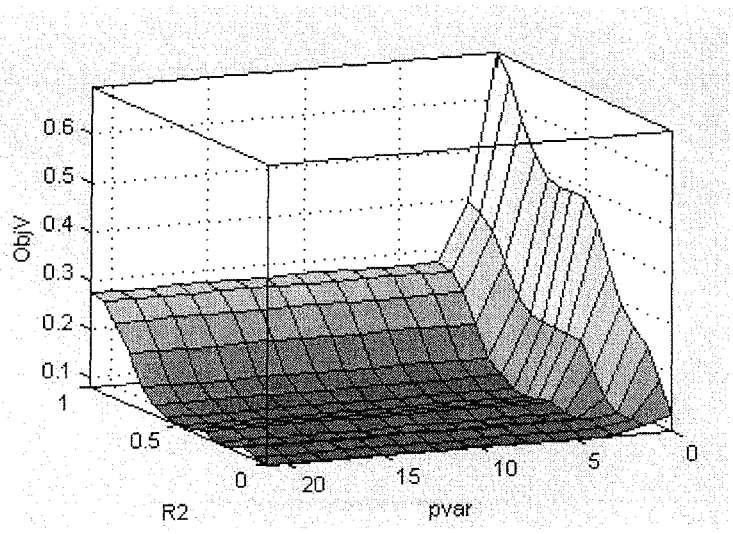

Figura 4. Andamento dell'indice *FIS* al variare di R^2 e pvar (costante N_var pari al suo valore medio).

Viceversa, ma sempre per lo stesso motivo, utilizzando il programma Backward (partendo quindi dal modello completo con l'eliminazione di volta in volta di una variabile) l' R^2 già alto non scende così velocemente da ridurre immediatamente la bontà complessiva del modello (anzi, aumenta durante le prime iterazioni). In tal caso, infatti, l'algoritmo effettua ben 8 (21 – 13 = 8) cicli prima di fermarsi. In Fig. 5 è possibile notare l'andamento dell'indice al variare di R^2 e di N_var (in questo caso si è posto costante il numero di parametri non significativi) e l'andamento dello stesso indice al variare di pvar e N_var (costante R^2).

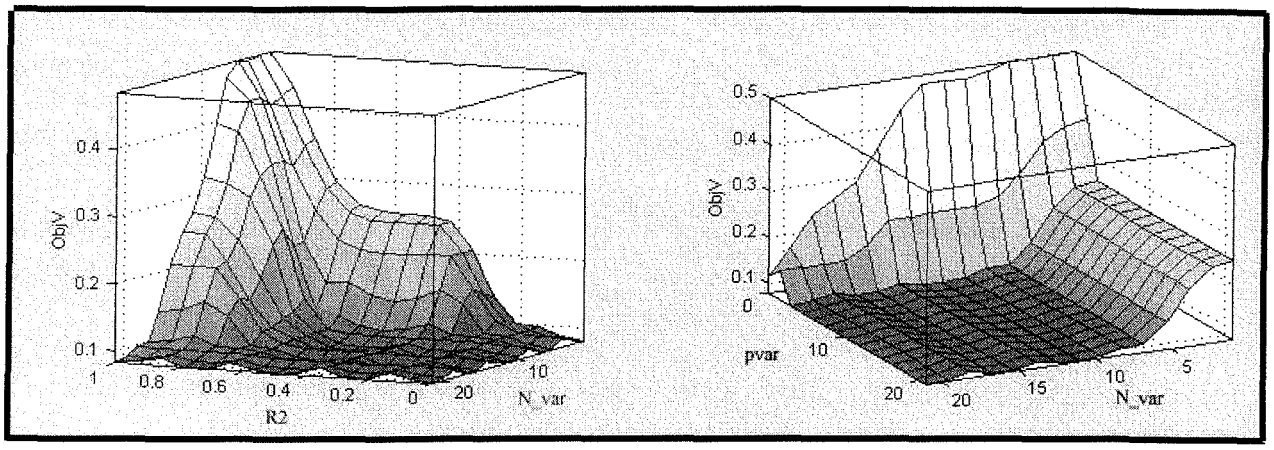

Figura 5. Andamento dell'indice *FIS* al variare di R^2 e N_var (costante pvar pari al suo valore medio) e al variare di pvar e N_var (costante R^2 pari al suo valore medio).

Non sono certamente i risultati ottenuti con queste due metodologie quelli più interessanti dal punto di vista tecnico in quanto gli stessi risultati si erano già ottenuti con altri indicatori (almeno per quanto concerne il programma "forward", ma risultati molto simili si erano ottenuti anche nel programma "Backward").

Merita invece un approfondimento il risultato ottenuto con l'Algoritmo Ibrido[4].

Così come la maggior parte delle applicazioni pratiche che utilizzano sistemi ibridi, anche questo tipo di algoritmo è stato appositamente studiato per trarre vantaggio dai rispettivi punti di forza delle due metodologie.

In questo caso si è così tentato di risolvere il duplice problema che sta alla base delle procedure classiche automatizzate: la necessità di avere un risultato che non dipendesse dalla sequenza di scelta delle variabili esplicative (a questo pone rimedio l'algoritmo genetico) e la necessità di evitare di costruire procedure completamente automatizzate che non tenessero in alcuna considerazione le valutazioni aprioristiche dell'utente (a ciò provvede la Logica Fuzzy).

I risultati così ottenuti sono molto buoni in quanto la dimensione del modello selezionato è bassa (due sole variabili selezionate entrambe significative), R^2_{test} è alto (0.772) e lo scarto quadratico medio previsivo è basso (4.12).

Da un punto di vista previsivo, forse è da preferire il risultato ottenuto con il programma "GALMS" applicato al criterio *AIC*, ma se l'obiettivo è quello di scegliere un modello che risulti semplice senza rinunciare alle buone capacità previsive, la scelta deve cadere su questo criterio *(FIS)*.

D'altronde, non si può universalmente stabilire se una delle due procedure è migliore dell'altra. La scelta dovrà effettuarsi in base alle necessità e agli scopi della ricerca.

È necessario fare un'ultima considerazione per quanto riguarda il tempo impiegato nell'esecuzione dei programmi. Lavorando con un processore Pentium II a 400 Mhz, il tempo computazionale necessario per eseguire tutti i 2.097.151 (2^{21}-1) possibili modelli sarebbe di circa 116 ore e 30 minuti (circa cinque modelli per ogni secondo).

Il programma "Backward" impiega per eseguire l'intero programma circa un minuto e 30 secondi, il programma "Forward" impiega 12 secondi mentre il programma "GALMS" impiega circa 23 secondi per eseguire una generazione di 100 individui (quindi complessivamente impiega circa 38 minuti).

È evidente quindi che lavorando con gli algoritmi genetici il tempo complessivamente impiegato è notevolmente maggiore rispetto a quello impiegato con gli altri tipi di programmi.

---

[4] Gli "Intelligent Hybrid System" o sistemi intelligenti ibridi [GoKh95] utilizzano più tecnologie intelligenti contemporaneamente. Nel caso della simulazione sopra esposta si è utilizzata la Logica Fuzzy contemporaneamente agli Algoritmi Genetici.

Questo può disincentivare l'utilizzo di tali tecniche che, comunque, complessivamente rispondono meglio alle attese del ricercatore essendo migliori in termini di efficienza e robustezza. Inoltre, portando a convergenza l'algoritmo, il tempo impiegato risulta ridotto di circa 200 volte rispetto a quello impiegato per la verifica di tutti i modelli (tempo richiesto applicando la strategia "All possible").

4. SELEZIONE DEL MODELLO PER SERIE STORICHE

Nel caso di serie temporali si ricorre a modelli lineari in cui l'eventuale componente non lineare viene introdotta come un termine perturbativo. In questo paragrafo viene affrontata la problematica della selezione di un modello per l' analisi di serie storiche utilizzando un approccio evolutivo analogo a quanto descritto nel paragrafo precedente. Lo strumento viene testato su serie storiche simulate.Tutti i modelli presentati vengono definiti esclusivamente come processi stocastici[5].
Se $X(t)$ è il valore del processo stocastico al tempo t ($t = 1,...,n$) un processo autoregressivo (AR) è definito dalla relazione che collega $X(t)$ ai valori precedenti:

$$X(t) = \Phi_1 X(t-1) + \Phi_2 X(t-2) + ... + \Phi_p X(t-p) + \varepsilon_t \qquad (12)$$

dove ε_t ($t = 1,...,n$) è una successione di variabili casuali incorrelate con media 0 e varianza costante σ_ε^2.
La scelta di un processo autoregressivo potrebbe perciò essere giustificata dall'esigenza di "spiegare" $X(t)$ collegandolo ai valori precedenti con un sistema di pesi $\Phi_1, \Phi_2, ..., \Phi_p$.
Tale procedura consente un rapido "aggiornamento" delle previsioni man mano che si aggiungono nuove osservazioni alla serie.
Fissato p, il problema di stimare in una serie storica il vettore $\Phi = (\Phi_1, \Phi_2 ..., \Phi_p)$ viene risolto con il metodo dei minimi quadrati (o con la stima di massima verosimiglianza) così come previsto nel modello lineare precedente. Il problema fondamentale, pertanto, è quello di fissare l'ordine p del processo *AR* che si ritiene ben adeguato a rappresentare la realtà dei fenomeni.
Per semplificare la notazione utilizzata indichiamo l'ordine del processo tra parentesi (per esempio un modello autoregressivo del quinto ordine verrà indicato con *AR(5)*) e indichiamo il numero dei parametri diversi da zero con un apice (se il modello presenta solo tre dei cinque parametri diversi da zero verrà indicato pertanto con la notazione $AR^3(5)$).
Il problema quindi non è soltanto quello di determinare l'ordine del processo, ma anche quello di determinare il numero di elementi del vettore $\Phi = (\Phi_1, \Phi_2 ..., \Phi_p)$ diversi da 0.
Pertanto il modello:

$$X(t) = \Phi_1 X(t-1) + \Phi_2 X(t-2) + \Phi_7 X(t-7) + \varepsilon_t \qquad (13)$$

identifica un processo autoregressivo del settimo ordine e viene così indicato: $AR^3(7)$.
Nonostante queste differenze di notazione, in termini di problematica generale, l'analisi del modello *AR* non si discosterà affatto da quella riguardante il modello lineare vista in precedenza. Si dovranno risolvere, quindi, sia il problema del metodo utilizzato per la selezione dei parametri, sia quello inerente il criterio da applicare (con il metodo scelto). In questa analisi si simuleranno delle

[5] Ogni variabile in un processo stocastico è una variabile casuale e le osservazioni evolvono nel tempo in base a determinate distribuzioni di probabilità. L'importanza dei processi stocastici in tali applicazioni è riconosciuta universalmente. Oggi, infatti, non è più consentito trattare con rigore una serie storica senza concepirla, nello stesso tempo, come "realizzazione" finita di un processo stocastico la cui struttura interna è e rimane fondamentalmente ignota al ricercatore.

serie storiche, quindi la struttura del modello sarà nota, e si tenterà mediante un algoritmo geneticamente evoluto di dare una risposta ai quesiti posti.
Il passo successivo è la simulazione di processi *ARMA* (autoregressivi – media mobile).
Una classe di processi stocastici può formarsi introducendo ai modelli AR un numero finito di ritardi nella media mobile. Il modello può essere così rappresentato:

$$X(t) = \Phi_1 X(t-1) + ... + \Phi_p X(t-p) + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q} \tag{14}$$

dove *(p,q)* rappresenta l'ordine del modello *ARMA*.
Anche per questa classe di modelli verrà utilizzata la notazione sopra presentata con le stesse osservazioni.
Il modello:

$$X(t) = \Phi_1 X(t-1) + \Phi_5 X(t-5) + \varepsilon_t + \theta_2 \varepsilon_{t-2} \tag{15}$$

viene indicato con la notazione $ARMA^{2,1}(5,2)$.

4. 1. L' ALGORITMO GATS

Vengono simulate alcune serie temporali (circa 200 complessivamente, ripartite tra processi *AR* e processi *ARMA*) generate da modelli il cui ordine viene stabilito casualmente in un intervallo [1,10], per quanto riguarda quelli *AR*, e negli intervalli ([1,10],[1,10]), per quanto riguarda i modelli *ARMA*.
Il termine di errore addizionato, generato casualmente da una distribuzione *N(0,1)*, viene moltiplicato per un coefficiente costante pari a 0.2.
Le serie così simulate sono ammesse poi alla fase successiva di analisi e stima solamente se il processo da esse generato risulta essere stazionario.
La struttura del programma ricalca quella vista per il modello lineare nel capitolo precedente con le poche variazioni dettate dalle diverse esigenze.
La codifica degli individui associati alle popolazioni viene quindi modificata soltanto nella sua lunghezza: la stringa associata al modello *AR* è rappresentata in Fig. 6, mentre quella associata al modello *ARMA* è rappresentata in Fig. 7.
La lunghezza delle stringhe corrisponde all'ordine massimo associato ai due tipi di modelli. Nella stringa associata al modello *ARMA* le prime 10 celle individuano i ritardi nella variabile esplicativa, mentre le celle dall'undicesima alla ventesima individuano i ritardi nella media mobile.

0	1	0	0	1	1	0	0	0	0

Figura 6. Rappresentazione dell'individuo utilizzato nell'algoritmo per indicare i modelli autoregressivi simulati.

Alla stringa rappresentata in Fig. 6 corrisponde pertanto il modello $AR^3(6)$:

$$X(t) = \Phi_2 X(t-2) + \Phi_5 X(t-5) + \Phi_6 X(t-6) + \varepsilon_t \tag{16}$$

AR										--						MA				
0	0	1	0	1	0	0	0	1	0		0	0	1	0	0	1	0	1	0	1

Figura 7. Rappresentazione dell'individuo utilizzato nell'algoritmo per indicare i modelli ARMA simulati..

Alla stringa rappresentata in Fig. 7 corrisponde, invece, il modello $ARMA^{3,4}(9,10)$:

$$X(t) = \Phi_3 X(t-3) + \Phi_5 X(t-5) + \Phi_9 X(t-9) + \varepsilon_t + \theta_3 \varepsilon_{t-3} + \theta_6 \varepsilon_{t-6} + \theta_8 \varepsilon_{t-8} + \theta_{10} \varepsilon_{t-10} \,. \tag{17}$$

Come detto, prima si sono simulati processi autoregressivi di ordine massimo 10, poi processi autoregressivi-media mobile di ordine massimo (10,10) e infine si è tentato di "trovare" la struttura del modello sottostante alla serie storica mediante l'uso di un Algoritmo Genetico.
La scelta del modello più appropriato è stata fatta utilizzando l'indicatore *AIC* in termini previsivi.
Per far ciò si è suddiviso il campione di dati generato (l'intero "data-set") in tre gruppi distinti:
Il training set contenente il 40% dell'intero campione, il validation set contenente un restante 40% e il test set contenente il restante 20%.
Il primo è stato applicato per la stima dei parametri del modello corrispondente all'individuo i-esimo della popolazione, sul secondo viene valutato il criterio di fitness dell' algoritmo genetico utilizzando i parametri del modello determinati sulla base dei valori del training set. Sul terzo si è calcolato l'*AIC* previsivo (una volta stimato il modello è stato infatti possibile prevedere i valori futuri).
Si è scelto di adottare l'*AIC* in quanto complessivamente si era mostrato un buon indicatore in tutte le situazioni in cui era stato utilizzato (per l'analisi critica dei vari indicatori implementabili come criteri di scelta del modello si vedano, infatti, i capitoli precedenti), e inoltre è un indice di raffronto molto utilizzato per la valutazione dei modelli statistici basati su serie temporali[6].
Per quanto riguarda la struttura parametrica dell'Algoritmo Genetico e il criterio di ottimizzazione, nulla è mutato rispetto al modello lineare esaminato in precedenza: uguali sono le dimensioni della popolazione, uguali i cicli iterativi massimi da compiere (MAXGEN), uguali le altre percentuali di mappatura.
Nelle tabelle 5 e 6 sono rappresentati i risultati ottenuti divisi in base all'ordine del modello statistico corrispondente.
Nella tabella 5 vengono evidenziati i risultati ottenuti con le simulazioni di modelli autoregressivi: nella prima colonna è indicato l'ordine del modello; nella seconda è riportata la percentuale di simulazioni generate con modelli di ordine indicato nella prima colonna che l'algoritmo ha correttamente previsto; nella terza colonna è invece indicata la percentuale di modelli che non è stata correttamente individuata, ma il programma ha selezionato un modello avente *AIC* minore a causa del minore S^2 (si tratta generalmente di modelli sovraparametrizzati); nella quarta colonna è indicata la percentuale di quei modelli non correttamente selezionati, ma che presentano un valore di *AIC* più basso dovuto al minor numero di parametri presenti nel modello da stimare; nella quinta colonna sono riportate le percentuali degli errori commessi dal programma in quanto l'algoritmo non riesce a convergere verso la soluzione; nell'ultima colonna è indicato il totale.

[6] Ricordiamo che l'*AIC* è definito come: $AIC(p) = n\log(S_p^2) + 2(p)$.

MODELLO	Modello correttamente identificato	Modello non identificato (modello proposto con S^2 minore)	Modello non identificato (modello proposto con AIC minore)	Modello non identificato (errore del programma)	TOTALE
AR(1)	100% (3)	-	-	-	100% (3)
AR(2)	100% (3)	-	-	-	100% (3)
AR(3)	50% (3)	50% (3)	-	-	100% (6)
AR(4)	100% (7)	-	-	-	100% (7)
AR(5)	90% (9)	10% (1)	-	-	100% (10)
AR(6)	86% (6)	-	14% (1)	-	100% (7)
AR(7)	100% (7)	-	-	-	100% (7)
AR(8)	64% (14)	18% (4)	18% (4)	-	100% (22)
AR(9)	92% (23)	4% (1)	4% (1)	-	100% (25)
AR(10)	88% (30)	3% (1)	6% (2)	3% (1)	100% (34)
TOTALE	85% (105)	8% (10)	6% (8)	1% (1)	100% (124)

Tabella 5. Risultati ottenuti sulle simulazioni di modelli AR..

Nelle colonne 2-6, in parentesi, sono riportati i valori assoluti delle simulazioni associate alle percentuali indicate.

Nella tabella 6 vengono presentati i risultati relativi alle simulazioni di modelli ARMA con le medesime indicazioni viste per la tabella 5.
Nelle tabelle 7 e 8, invece, vengono mostrati gli stessi risultati delle tabelle 5 e 6 suddivisi però in base alla dimensione del modello generatore della serie storica corrispondente.
Nella tabella 7 rientrano i modelli *AR* mentre nella tabella 8 quelli *ARMA*.
Com'è possibile osservare dalle tabelle presentate, i modelli selezionati non coincidono sempre con il modello utilizzato per la simulazione in quanto l'algoritmo converge verso una soluzione ottimale differente.
La causa di questo inconveniente è duplice: i modelli sovraparametrizzati a volte abbassano l' S_p^2 più di quanto si aumenti il termine 2p, mentre modelli sottoparametrizzati a volte abbassano il termine 2p più di quanto possa crescere l' S_p^2. In secondo luogo, tale indicatore viene calcolato sulle previsioni effettuate e non sul campione di dati su cui vengono stimati i parametri. Queste previsioni non possono includere un termine di errore così come capita sui dati di confronto delle stesse: la conseguenza è che, se la componente di errore è molto "forte", l'algoritmo può stimare un modello differente che però in termini di *AIC* previsivo risulta migliore.
Quindi tutti i modelli stimati non correttamente, riportati in colonna tre e quattro non devono considerarsi, pertanto, errori del programma o difetti dell'algoritmo che non riesce a convergere, ma problemi connessi all'uso di indicatori non sempre efficienti o all'errore stocastico aggiunto alla serie.
Per meglio comprendere quanto possa essere importante questa affermazione, vengono di seguito presentati alcuni esempi di modelli simulati , stimati e previsti dal programma.

MODELLO	Modello correttamente identificato	Modello non identificato (modello proposto con S^2 minore)	Modello non identificato (modello proposto con AIC minore)	Modello non identificato (errore del programma)	TOTALE
ARMA(1,1)	100% (3)	-	-	-	100% (3)
ARMA(2,2)	25% (1)	50% (2)	25% (1)	-	100% (4)
ARMA(3,3)	20% (2)	40% (4)	40% (4)	-	100% (10)
ARMA(4,4)	-	100% (4)	-	-	100% (4)
ARMA(5,5)	-	100% (4)	-	-	100% (4)
ARMA(6,6)	-	100% (5)	-	-	100% (5)
ARMA(7,7)	-	50% (3)	50% (3)	-	100% (6)
ARMA(8,8)	-	100% (4)	-	-	100% (4)
ARMA(9,9)	-	100% (4)	-	-	100% (4)
ARMA(10,10)	-	100% (4)	-	-	100% (4)
ARMA(1,2)	100% (4)	-	-	-	100% (4)
ARMA(1,3)	80% (4)	20% (1)	-	-	100% (5)
ARMA(1,4)	50% (3)	33% (2)	17% (1)	-	100% (6)
ARMA(1,5)	50% (2)	50% (2)	-	-	100% (4)
ARMA(2,1)	100% (6)	-	-	-	100% (6)
ARMA(2,3)	60% (3)	20% (1)	20% (1)	-	100% (5)
ARMA(2,4)	-	100% (8)	-	-	100% (8)
ARMA(2,5)	25% (2)	75% (6)	-	-	100% (8)
TOTALE	32% (30)	57% (54)	11% (10)	-	100% (94)

Tabella 6. Risultati ottenuti sulle simulazioni di modelli ARMA.

MODELLO	Modello correttamente identificato	Modello non identificato (modello proposto con S^2 minore)	Modello non identificato (modello proposto con AIC minore)	Modello non identificato (errore del programma)	TOTALE
AR^1	95% (19)	5% (1)	-	-	100% (20)
AR^2	75% (15)	10% (2)	15% (3)	-	100% (20)
AR^3	90% (18)	5% (1)	5% (1)	-	100% (20)
AR^4	90% (18)	5% (1)	5% (1)	-	100% (20)
AR^5	80% (16)	15% (3)	5% (1)	-	100% (20)
AR^6	79% (19)	13% (3)	4% (1)	4% (1)	100% (24)
TOTALE	85% (105)	8% (11)	6% (7)	1% (1)	100% (124)

Tabella 7. Risultati ottenuti sulle simulazioni di modelli AR.

MODELLO	Modello correttamente identificato	Modello non identificato (modello proposto con S^2 minore)	Modello non identificato (modello proposto con AIC minore)	Modello non identificato (errore del programma)	TOTALE
ARMA(1,1)	60% (9)	40% (6)	-	-	100% (15)
ARMA(1,2)	47% (7)	47% (7)	6% (1)	-	100% (15)
ARMA(2,1)	40% (6)	47% (7)	13% (2)	-	100% (15)
ARMA(2,2)	33% (5)	47% (7)	20% (3)	-	100% (15)
ARMA(3,1)	20% (3)	67% (10)	13% (2)	-	100% (15)
ARMA(3,2)	-	89% (17)	11% (2)	-	100% (19)
TOTALE	32% (30)	57% (54)	11% (10)	-	100% (94)

Tabella 8. Risultati ottenuti sulle simulazioni di modelli ARMA.

La Fig. 8 mostra con il colore rosso il test-set di una serie storica generata dal seguente modello:

$$X(t) = 0.3X(t-3) - 0.35X(t-6) + 0.4X(t-8) - 0.25X(t-9) - 0.3X(t-10) + \varepsilon_t \tag{18}$$

con il colore blu, invece, vengono evidenziate le previsioni ottenute con il "migliore" dei modelli stimati:

$$X(t) = 0.261X(t-3) - 0.376X(t-6) + 0.304X(t-8) - 0.281X(t-9) - 0.209X(t-10) \tag{19}$$

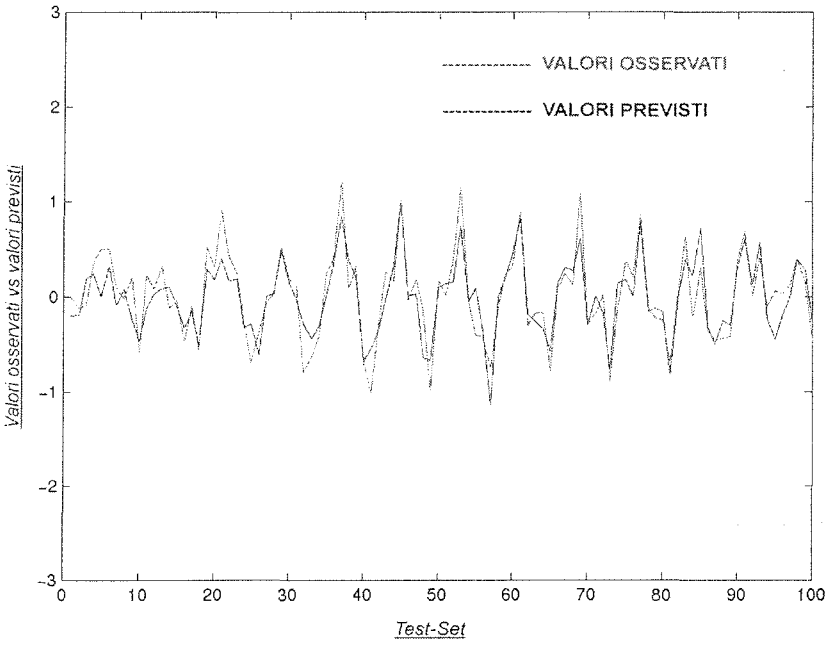

Figura 8. Previsioni vs valori reali (osservati) di una serie storia generata da un modello autoregressivo del decimo ordine.

Come si può notare l'algoritmo converge correttamente verso il modello generatore della serie storica in quanto questo minimizza l'*AIC* previsivo. La differenza riscontrata nei coefficienti parametrici è causata dal termine di errore del modello generatore.
In questo caso l'algoritmo identifica correttamente il modello: $AR^5(10)$.

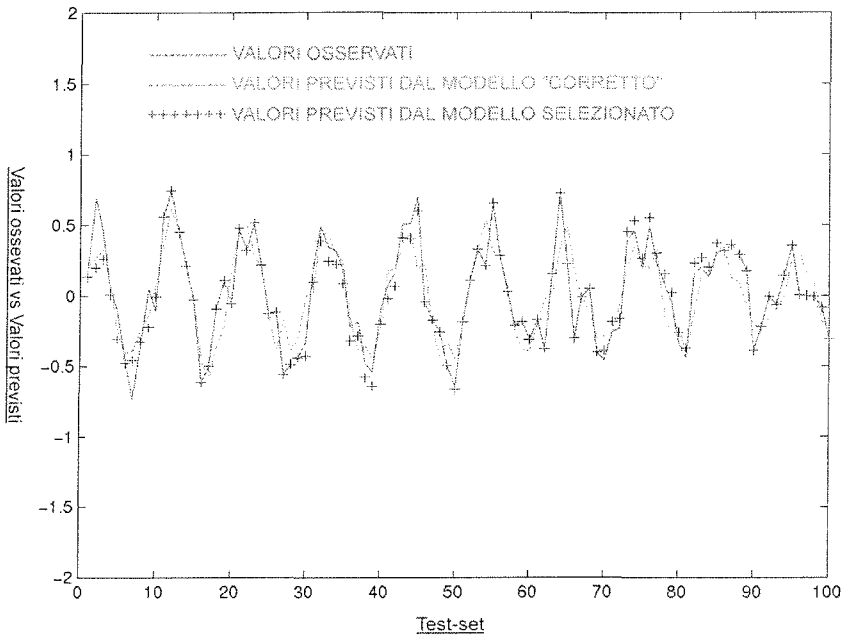

Figura 9. Previsioni (stimate = blu, modello corretto = verde) vs valori osservati (rosso) di una serie storia generata da un modello autoregressivo del nono ordine.

In Fig. 9 è invece rappresentata la serie storica (in rosso) generata con il modello $AR^5(9)$:

(20) $$X(t) = -0.4X(t-1) + 0.3X(t-2) + 0.4X(t-5) + 0.3X(t-7) + 0.3X(t-9) + \varepsilon_t$$

In blu (" + ") sono evidenziate le previsioni stimate con il modello selezionato, $AR^4(9)$:

(21) $$X(t) = -0.359X(t-1) + 0.343X(t-5) + 0.137X(t-7) + 0.348X(t-9)$$

In questo caso l'algoritmo non converge correttamente verso il modello generatore della serie storica poiché non è quello che minimizza l'*AIC* previsivo: l'algoritmo seleziona il modello "migliore" dal punto di vista previsivo.
Nella Fig. 9 è possibile vedere anche le previsioni stimate se il modello fosse stato correttamente individuato (in colore verde). Le previsioni risulterebbero notevolmente peggiorate se fosse stato selezionato questo modello.
La causa di questo, come detto, può essere imputata al termine di errore (o eventualmente all'indicatore utilizzato che, in questo caso, sottoparametrizzerebbe il modello).
Se per i modelli *AR* il problema non è molto rilevante, così come si evidenzia nella tabella 5 (in cui l'85% dei modelli simulati è correttamente identificato), per i modelli *ARMA* non si può dire lo stesso. Il problema non è affatto trascurabile: spesso (in quasi il 70% dei casi) l'algoritmo tende a convergere verso il minimo *AIC* che non corrisponde a quello del modello utilizzato per la simulazione (vedi tabella 5).

Le Figg. 10 e 11 mostrano le previsioni stimate sulle serie storiche generate da un modello *ARMA*.
In particolare la Fig. 10 mostra con il colore rosso il test-set di una serie storica generata con il seguente modello:

(22) $$X(t) = 0.3X(t-3) - 0.4X(t-5) + 0.2X(t-7) + \varepsilon_t + 0.11\varepsilon_{t-2} - 0.15\varepsilon_{t-6}$$

mentre con il colore blu sono evidenziate le previsioni ottenute con il migliore dei modelli stimati:

(23) $$X(t) = 0.27X(t-3) - 0.4X(t-5) + 0.22X(t-7) + \varepsilon_t + 0.1\varepsilon_{t-2} - 0.12\varepsilon_{t-6}$$

L'algoritmo converge correttamente verso il modello generatore della serie storica in quanto questo minimizza l'*AIC* previsivo.

In Fig. 11 è invece rappresentato il test set di una sere storica (in colore rosso) generata dal seguente modello:

(24) $$X(t) = -0.1X(t-8) + \varepsilon_t + 0.35\varepsilon_{t-1}$$

In verde sono evidenziate le previsioni ottenute con il modello previsivo selezionato dall'algoritmo:

(25) $$X(t) = -0.38X(t-1) + 0.05X(t-8) + \varepsilon_t + 0.06\varepsilon_{t-3}$$

In blu invece vengono mostrate le previsioni ottenibili se l'algoritmo avesse correttamente individuato il modello generatore della serie storica.
Anche in questo caso l'algoritmo seleziona un modello migliore per quanto riguarda l'*AIC* previsivo.

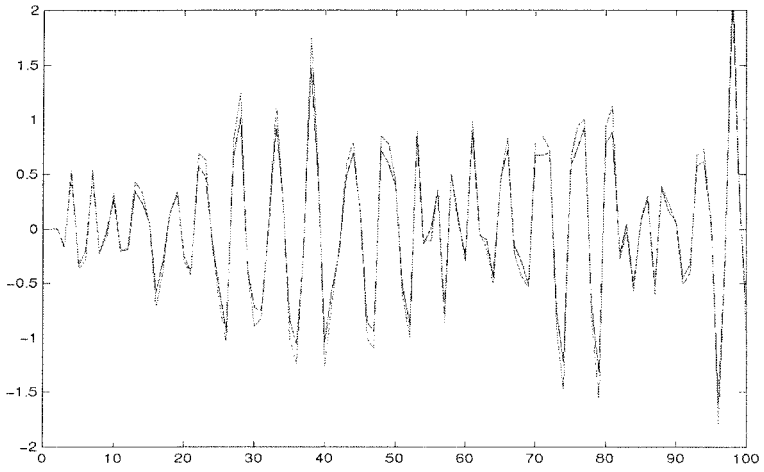

Figura 10. Previsioni (blu) vs valori osservati (rosso) di una serie storia generata da un modello ARMA^{3,2}(7,6).

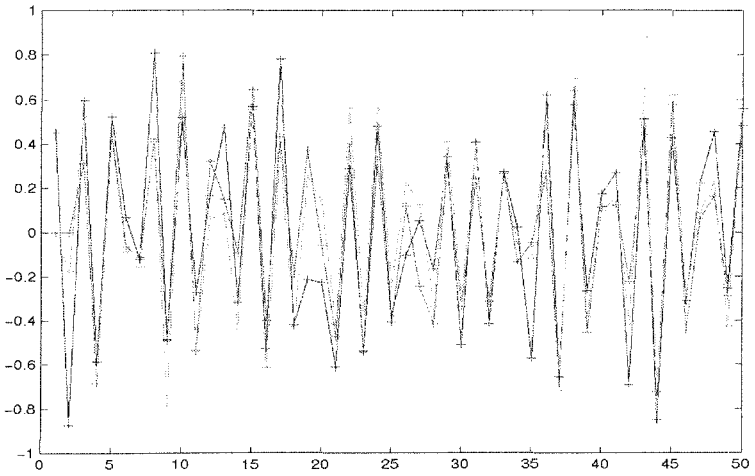

Figura 11. Previsioni (stimate = verde, modello corretto = blu) vs valori osservati (rosso) di una serie storia generata da un modello ARMA.

Ulteriore approfondimento può effettuarsi esaminando le matrici di confusione dei risultati ottenuti utilizzando diversi criteri di selezione: AIC, BIC e SIC.

AR	*1*	*2*	*3*	*4*	*5*	*6*
1	95 %	15 %				
2	5 %	75 %	5 %			
3		5 %	90 %	5 %	5 %	
4		5 %	5 %	90 %		4 %
5				5 %	80 %	5 % err.
6					10 %	79 %
7					5 %	8 %
8						
9						4 %
10						

Tabella 9. Indicatore utilizzato : AIC (esperimenti complessivi circa 170).

AR	*1*	*2*	*3*	*4*	*5*	*6*
1	100 %	50 %	25 %			
2		50 %	20 %		10 %	
3			50 %	66 %	20 %	20 %
4			5 % err.	33 %	30 %	30 %
5					40 %	20 %
6						40 %
7						
8						
9						
10						

Tabella 10. Indicatore utilizzato : SIC (esperimenti complessivi 60).

AR	*1*	*2*	*3*	*4*	*5*	*6*
1						
2		100 %	20+10err. %			
3			70 %	33 %	20 %	10 %
4	20 %			66 %	50 %	30 %
5	20 %				10 %	30 %
6					20 %	10 %
7	20 %					20 %
8	40 %					
9						
10						

Tabella 11. Indicatore utilizzato : BIC (esperimenti complessivi circa 60).

Nelle tre tabelle, in colonna sono indicati il tipo di modello AR utilizzato per la simulazione dei dati (dove per tipo si intende il numero di parametri che caratterizzano il modello indipendentemente dalla posizione occupante all'interno dei 10 campi utilizzabili). In riga, invece, è rappresentato il tipo di modello identificato dal programma GATS.
"Err." indica la percentuale di modelli non stimata correttamente. Cioè le posizioni dei coefficienti nel modello non sono corrette indipendentemente dal numero degli stessi.

Utilizzando come indicatore il SIC si ha un a netta tendenza alla sottoparametrizzazione (ad eccezione degli AR(1), ovviamente) che si accentua sempre più aumentando il numero di parametri utilizzati nel modello di simulazione.
I risultati ottenuti con l'indicatore BIC, invece, mostrano un sovraparemetrizzazione quando l'ordine del modello utilizzato è basso (come l'AR(1)) mentre mostrano una sottoparametrizzazione quando il modello utilizzato per la simulazione si complica.
Molto più stabili verso il modello effettivamente utilizzato sono i risultati ottenuti utilizzando l'AIC. In questo caso si ha una tendenza variabile in ugual misura a sovraparametrizzare o sottoparametrizzare il modello. Soltanto aumentando il numero dei parametri (AR(5) e AR(6)) si ha un maggiore tendenza alla sovraparametrizzazione rispetto alla sottoparametrizzazione.
Complessivamente il numero di modelli correttamente individuato risulta notevolmente maggiore utilizzando l'AIC (oltre l'80 %) rispetto a quanto accade utilizzando gli altri due indicatori (circa il 50 %).
La percentuale di modelli stimati in cui il valore dell'indicatore (AIC, BIC, SIC) risulta inferiore a quello del modello utilizzato per la simulazione (il modello non è correttamente identificato ma la procedura porta alla scelta di un modello comunque migliore in termini dell'indicatore utilizzato rispetto a quello utilizzato per la simulazione) è superiore al 99%.

5. APPLICAZIONE DELLE TECNICHE EVOLUTIVE DI SELEZIONE DEL MODELLO

In questo capitolo verranno presentati due esempi pratici di implementazione di alcuni degli strumenti esaminati precedentemente. Verranno analizzate le problematiche da affrontare nella scelta del modello migliore per gli scopi che l'utente si propone in due casi tra loro molto differenti.

Il primo esempio si riferisce all'individuazione di un modello operativo previsionale del livello di marea della laguna veneta. I dati si riferiscono all' anno 1984. Il secondo riguarda la selezione di un modello che spieghi la durata in mesi del livello di disoccupazione rilevato in Emilia Romagna (sulla base dei dati raccolti dall'ISTAT e riferiti al periodo primo trimestre 1993 – primo trimestre 1995).

Non vengono proposte soluzioni efficienti e ottimali per tutti i problemi che si devono affrontare, ma solo alcuni aspetti di questi e come utilizzare gli strumenti esaminati per risolverli. Si tratta

pertanto di un'analisi volta a mostrare come le diverse tecniche strumentali perfezionate in questa trattazione possano interagire quando la situazione che si presenta ne permette l'utilizzo.

5. 1. MODELLO PREVISIONALE DELLE MAREE DI VENEZIA

Il singolare andamento delle maree nel mare Adriatico ha, da molti anni, attratto fisici e meteorologi impegnandoli in più direzioni per la realizzazione di schemi numerici in grado di interpretare il più fedelmente possibile il livello di marea. Elemento di particolare attrazione si è dimostrato il fenomeno dell'acqua alta che interessa principalmente le lagune venete, soprattutto quella di Venezia con l'allagamento del centro storico e delle isole. Questo evento provoca, da molti anni, danni rilevanti agli operatori economici veneziani, ai trasporti pubblici e privati, alla viabilità pedonale, all'attività portuale e da ultimo, ma non meno importante, causa condizioni di vita insalubri.

Gli studi per la previsione del livello di marea sono iniziati con la formulazione fisico-matematica del fenomeno mareale impostando le relazioni matematiche (equazioni dell'idrodinamica) dell'interazione aria-mare attraverso le condizioni iniziali e al contorno. Si ottiene così la distribuzione nel tempo e nello spazio di alcune variabili quali per esempio la velocità di corrente e il livello marino. Questo metodo, nonostante la buona affidabilità nella previsione, ha dimostrato ben presto grossi limiti, come la necessità di conoscere in tempo reale alcuni parametri meteomarini o la necessità di un potente mezzo di calcolo per ottenere velocemente i risultati. Negli ultimi anni quindi si sono seguite strade alternative, apparentemente più superficiali, fondate non più sulla formulazione idrodinamica (metodo deterministico) ma su schemi per lo più empirici i cui risultati si sono rivelati interessanti grazie all'applicazione delle teorie statistiche (metodo empirico-statistico). L'esistenza delle sesse (oscillazioni libere del mare) con un periodo di 22 ore per la frequenza fondamentale, suggerisce l'autoregressione, o l'uso dei livelli osservati nelle ore precedenti come predittori. In altre parole, questo significa che se l'Adriatico fosse già in oscillazione, se non ci fossero nuovi disturbi e se non ci fosse smorzamento, il suo livello risulterebbe lo stesso di 22 ore prima. A questo si aggiunge il fattore meteorologico: sia la pressione atmosferica (effetto barometro) che il vento col suo trascinamento hanno effetti notevoli; poiché il vento dipende soprattutto dalla pressione l' inclusione dei dati di pressione tiene conto, in prima approssimazione, anche dei possibili effetti del vento.

Anche la dipendenza non lineare tra la velocità del vento e il suo impatto sul mare può essere approssimata con la linearità, almeno come tentativo tuttavia la validità di queste approssimazioni si potrà giudicare solo dalla bontà dei risultati.

Una parte piuttosto consistente del fenomeno mareale viene generata dalle forze gravitazionali e poiché queste sono attualmente ben definite, l'analisi tecnica verrà basata soltanto sulla parte non astronomica del fenomeno.

Il modello è strutturato su un massimo di 182 predittori: : si cerca tra tutti i possibili 2^{182}-1 modelli quello "migliore" per ogni anticipo previsivo effettuato (1, 3, 6, 9, 12, 15 e 24 ore).

L'insieme dei predittori è così suddiviso: 132 livelli di marea osservata a Venezia, all'indietro nel tempo, da una certa ora T all'ora T-131 (cosiddetti parametri marini), 50 valori di pressione atmosferica osservati a Bari, Falconara, Genova, Pesaro, Pescara, Ravenna, Rimini, Teramo, Trieste, Venezia, ogni 3 ore dal tempo T al tempo T-12 (cosiddetti parametri metereologici).

Lo schema si presenta così:

$$h_{T+\tau} = \sum_{i=0}^{I} \left(a_i^{(\tau)} \times h_{T-i}\right) + \sum_{j=1}^{J} \sum_{k=0}^{K} b_{j,k}^{(\tau)} \times p_{(j,T-3k)} \qquad (26)$$

dove:

- T = tempo in cui viene effettuata la previsione;

- τ = anticipo previsivo in ore che si vuole ottenere;
- $h_{T+\tau}$ = livello di marea previsto per il tempo T più il numero di ore di anticipo τ ;
- $a_i^{(\tau)}$ = coefficienti relativi alla marea a Venezia dello schema regressivo con $i=0,...,131$ (massimo 132 coefficienti per ogni anticipo di previsione τ);
- h_{T-i} = livelli di marea osservati a Venezia al tempo T-i (dal tempo T al tempo T-131 a seconda dei coefficienti selezionati);
- $b_{j,k}^{(\tau)}$ = coefficienti dello schema regressivo per le pressioni atmosferiche registrate nelle stazioni j ($j=1,2,...,10$; nell'ordine Bari, Falconara, Genova, Pesaro, Pescara, Ravenna, Rimini, Teramo, Trieste e Venezia) relativi alle 5 osservazioni triorarie k ($k=0,1,2,3,4$) di pressione atmosferica (sono state utilizzate massimo 5 osservazioni triorarie di pressione dal tempo T al tempo T-12 per ognuna delle 10 stazioni metereologiche);
- $p_{j,T-3k}$ = pressioni atmosferiche registrate a Bari, Falconara, Genova, Pesaro, Pescara, Ravenna, Rimini, Teramo, Trieste e Venezia (indice j) nelle ore $T, T-3, ..., T-12$.

I coefficienti (a_i e $b_{j,k}$) sono quindi, complessivamente, 182.

Per ogni anticipo τ la specificazione del modello "migliore" avviene mediante l'utilizzo di Algoritmi Genetici (utilizzando l' algoritmo di selezione del modello lineare "GALMS" con l'introduzione della componente autoregressa). Il programma seleziona tra i 182 predittori quelli che specificheranno il modello "migliore". Come criterio di fitness e' stato utilizzato l' AIC calcolato sulle previsioni.

I coefficienti vengono calcolati sulla base dei dati di marea e pressione osservati nell'anno 1984 tra l'1/1/84 e il 19/10/84 (training set) secondo i metodi statistici classici (stima OLS). In fase previsionale lo schema proposto è stato utilizzato per l'anno 1984 tra il 19/10/84 e il 31/12/84 (test set). A livello sperimentale è stato protratto il valore di τ fino a 24 ore (assumendo, come detto sopra, i valori 1, 3, 6, 9, 12, 15, 24).

La codifica dell'algoritmo prevede pertanto l'utilizzo di una stringa costituita da 182 celle, ognuna delle quali può assumere valore 0 o 1 (0 indicherà l'esclusione di quel predittore dal modello, 1 indicherà invece l'inclusione).

Si è provveduto quindi a confrontare, per tutti i casi di alta marea (in cui il livello di marea astronomica sommato alla componente metereologica risulta superiore ai 100 centimetri di altezza) verificatisi dal 23 settembre 1984 al 31 dicembre 1984, l'indicazione del modello previsivo con l'effettivo azionamento o meno delle sirene per valori di marea superiori a 110 centimetri.

Nelle Fig. 12-14 vengono presentati i risultati previsivi generali ottenuti con un'anticipazione previsiva di un'ora.

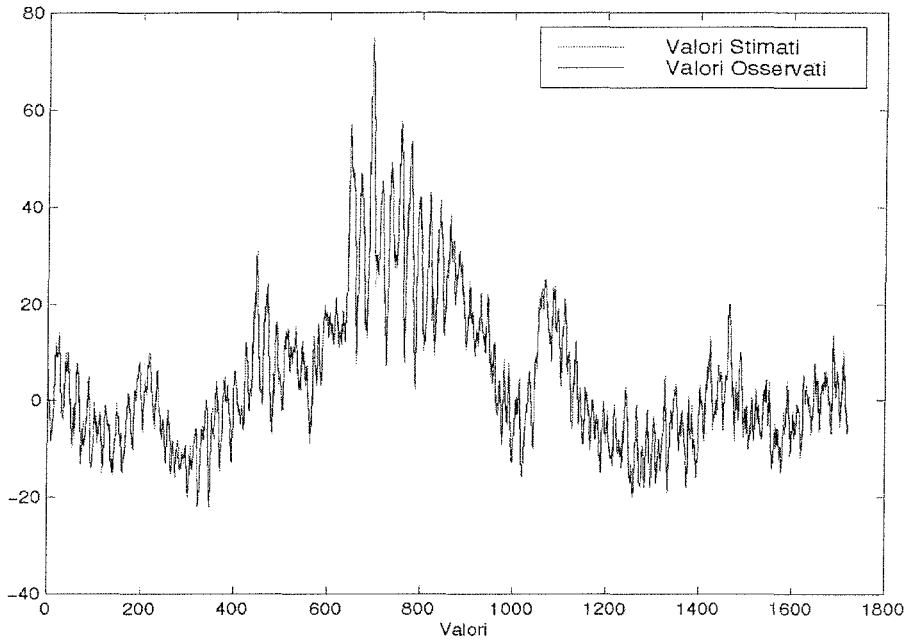

Figura 12. Valori previsti vs valori osservati con un'anticipazione previsiva di un'ora ottenuta con modello lineare (previsioni effettuate dal 19/10/84 ore 04.00 al 31/12/84 ore 24.00).

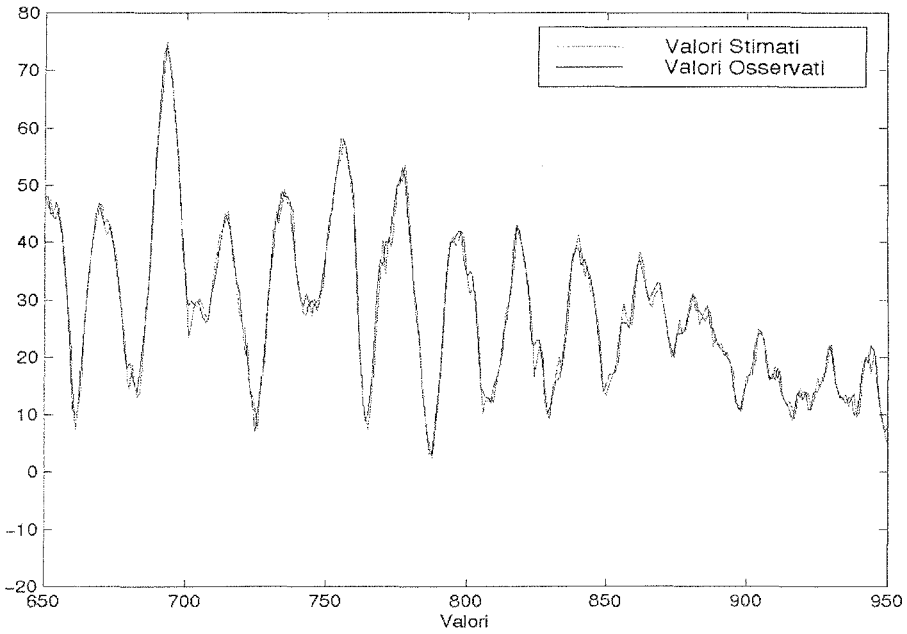

Figura 13. Valori previsti vs valori osservati con un'anticipazione previsiva di un'ora (previsioni effettuate dal 15/11/84 ore 5.00 al 27/11/84 ore 17.00) ottenuta con modello lineare.

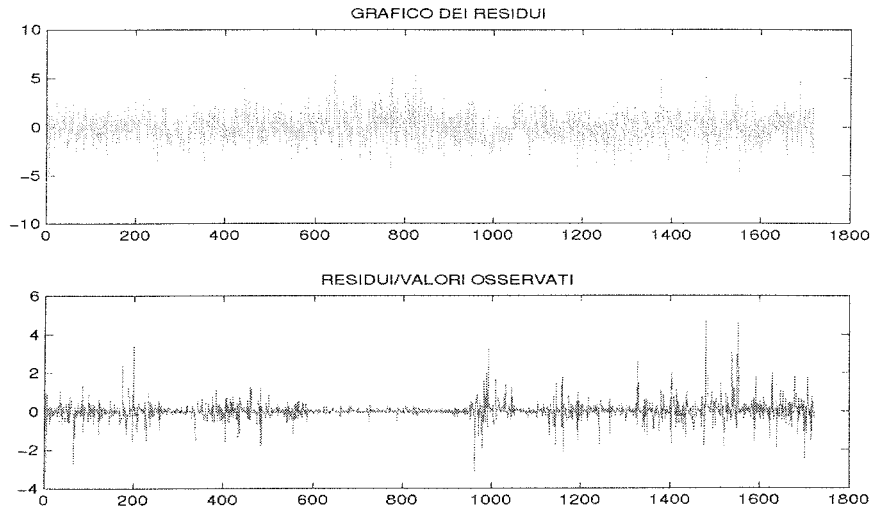

Figura 14. Residui ottenuti sulla previsione a un'ora (previsioni effettuate dal 19/10/84 ore 4.00 al 31/12/84 ore 24.00) ottenuta con modello lineare.

Sono stati successivamente presi in esame, per tutti i casi di alta marea (almeno 100 cm.) verificatisi dal 23 settembre 1984 al 31 dicembre 1984, il livello previsto dai modelli sopra esposti e il relativo valore di raffronto osservato.

Nelle tabelle che seguono, nella prima riga sono elencati i valori degli anticipi previsti rispetto al giorno e all'ora riportati sopra ad ognuna di esse; nella seconda riga vengono mostrati i livelli di marea previsti per ogni anticipo previsivo; nella terza riga vengono valutati, per ogni anticipazione oraria, se le sirene di allarme, che a Venezia devono suonare ogni volta che si prevede un livello di marea superiore a 110 cm., sarebbero state azionate correttamente.

I risultati mostrano una buona capacità previsiva del modello e spesso si presentano migliori dei sistemi di riferimento concretamente utilizzati a Venezia. Non sempre però i valori previsti si avvicinano ai valori osservati, soprattutto quando l'anticipo supera le 3 ore. Molto indicative sono, infatti, le tabelle e le figure sugli andamenti previsivi che mostrano come il modello usato rilevi i picchi di minimo e di massimo senza ritardi temporali (così come avveniva, invece, con le previsioni di tipo finanziario). In particolare le due tabelle relative al giorno 23/09/84 evidenziano questa caratteristica mostrando un livello di marea inferiore ai 110 cm con ben 24 ore di anticipo (a Venezia, invece, sono suonate le sirene di allarme poiché il modello previsivo utilizzato ha fallito).

Giorno: 23/09/84 ore 20.00. Valore di raffronto osservato: 103 cm

Ore prev.	1	3	6	9	12	15	24
Valore	102.5 cm	102.9 cm	97.6 cm	95.9 cm	86.1 cm	87.5 cm	75.6 cm
Confronto	Corretto	Corretto	Corretto	Corretto	Corretto	Corretto	Corretto

Giorno: 23/09/84 ore 22.00. Valore di raffronto osservato: 105 cm

Ore prev.	1	3	6	9	12	15	24
Valore	107.0 cm	104.4 cm	96.7 cm	96.2 cm	90.3 cm	86.7 cm	77.9 cm
Confronto	Corretto	Corretto	Corretto	Corretto	Corretto	Corretto	Corretto

Giorno: 20/11/84 ore 9.00. Valore di raffronto osservato: 115 cm

Ore prev.	1	3	6	9	12	15	24
Valore	112.2 cm	106.5 cm	101.9 cm	98.9 cm	97.9 cm	102.3 cm	105.4 cm
Confronto	Corretto	Non corretto	Non corretto	Non corretto	Non corretto	Non corretto	Non corretto

Giorno: 21/11/84 ore 9.00. Valore di raffronto osservato: 119 cm

Ore prev.	1	3	6	9	12	15	24
Valore	117.3 cm	119.9 cm	113.9 cm	119.1 cm	119.6 cm	122.5 cm	116.5 cm
Confronto	Corretto	Corretto	Corretto	Corretto	Corretto	Corretto	Corretto

Giorno: 22/11/84 ore 9.00. Valore di raffronto osservato: 111 cm

Ore prev.	1	3	6	9	12	15	24
Valore	111.1 cm	113.0 cm	103.9 cm	108.1 cm	106.6 cm	107.9 cm	104.8 cm
Confronto	Corretto	Corretto	Non corretto	Non corretto	Non corretto	Non corretto	Non corretto

Giorno: 23/11/84 ore 10.00. Valore di raffronto osservato: 103 cm

Ore prev.	1	3	6	9	12	15	24
Valore	102.8 cm	101.9 cm	99.4 cm	101.8 cm	90.2 cm	94.9 cm	98.2 cm
Confronto	Corretto	Corretto	Corretto	Corretto	Corretto	Corretto	Corretto

Giorno: 24/11/84 ore 10.00. Valore di raffronto osservato: 100 cm

Ore prev.	1	3	6	9	12	15	24
Valore	100.0 cm	97.8 cm	92.9 cm	91.7 cm	92.0 cm	87.1 cm	89.0 cm
Confronto	Corretto	Corretto	Corretto	Corretto	Corretto	Corretto	Corretto

Tabella 12. Previsione del livello di marea a diverse anticipazioni orarie: i risultati.

5. 2. ANALISI SUL LIVELLO DI DISOCCUPAZIONE IN EMILIA ROMAGNA

In questo paragrafo vengono prese in esame le strategie implementate per la determinazione dei parametri di un modello statistico impiegato per spiegare correttamente il livello di disoccupazione mensile rilevato in Emilia Romagna.

I valori osservati e i parametri utilizzabili si riferiscono a un'indagine condotta da M. Lalla e F. Pattarin [LaPa99] su dati forniti dall'Istituto Centrale di Statistica (ISTAT) per la regione Emilia Romagna, dal primo trimestre 1993 al primo trimestre 1995. L'indagine è stata effettuata su un campione di 655 soggetti in cerca di occupazione, con un'età maggiore o uguale a 15 anni.

Il campione è stato suddiviso in tre sottogruppi: persone che hanno perso l'occupazione precedente per licenziamento, dimissioni o scadenza del contratto di lavoro a termine (*disoccupati*); persone che entrano per la prima volta nel mercato del lavoro e non hanno mai cercato in precedenza un'attività lavorativa (*nuovi ingressi*); persone che hanno smesso un'attività lavorativa e per un certo periodo di tempo non hanno più cercato un impiego uscendo così dalle forze di lavoro e ora hanno ricominciato a cercare un lavoro (*rientri*).

Per ogni componente della famiglia degli intervistati in età lavorativa sono stati rilevati la condizione professionale, notizie sull'attività stessa, durata della disoccupazione in corso e caratteristiche individuali che potrebbero influenzarla. Queste informazioni riguardano dati di carattere anagrafico, grado di istruzione dell'intervistato, componenti della famiglia, attività e periodi di disoccupazione precedenti, disponibilità, aspettative e abilità professionali, tipo di occupazione trovata dopo un periodo di disoccupazione e azioni compiute durante il periodo di disoccupazione.

Ponendo y uguale alla durata del periodo di disoccupazione espresso in mesi, il modello lineare può così rappresentarsi:

$$y = \sum_{i=1}^{n} a_i x_i \quad ; \tag{27}$$

dove n corrisponde al numero dei parametri relativi a ogni domanda fatta agli intervistati, in questo caso il massimo è 53; a_i rappresenta il coefficiente stimato per ogni variabile esplicativa utilizzata (x_i).

Pertanto x_i costituisce il vettore riga dei regressori utilizzati relativi alle caratteristiche personali, famigliari e del mercato del lavoro attuale sopra accennate.

La scelta dei regressori è stata effettuata utilizzando i programmi "Backward", "Forward" e quello di selezione genetica ("GALMS") analizzati nel quarto capitolo.

L'utilizzo di un modello lineare per rappresentare la durata della disoccupazione richiede l'inserimento dell'intero set di dati nel training set. Il motivo per cui è stata presa questa decisione è da ricercare nell'uso di alcune variabili "dummy" tra quelle esplicative. L'uso di questo tipo di variabili, infatti, potrebbe causare l'inserimento dell'intero campione rappresentativo della variabile stessa nel test set ponendo non pochi problemi nel momento in cui si devono stimare i parametri del modello. Inoltre, l'alta variabilità della y (con valori compresi tra 0.1 e 100.5) non perdona un eventuale sbilanciamento del numero di dati rappresentativi per quelle "dummy" tra il training e il test set[7].

I risultati ottenuti sono mostrati in tabella 13. Nella prima colonna, "metodo" indica il tipo di programma utilizzato per ottenere i risultati, nella seconda colonna "criterio" indica quale indice di riferimento è stato utilizzato, nella terza colonna è evidenziato il numero di variabili scelte dal programma per quel tipo di criterio (dimensione del modello ottimo), nella quarta colonna viene riportato il valore della stima dello scarto quadratico medio e nell'ultima colonna è stato inserito l' R^2 .

[7] I problemi aumenterebbero maggiormente se si prevedesse il campionamento dei dati in tre gruppi: training, validation e test set.

METODO	CRITERIO	DIMENSIONE	SQM	R^2
Backward	AIC	17	12.6916	0.1321
Backward	*BIC*	14	12.8349	0.1122
Backward	*HAN*	14	12.8349	0.1122
Backward	*SIC*	14	12.8349	0.1122
Backward	*p-value*	11	12.9091	0.1021
Forward	*AIC*	18	12.6890	0.1323
Forward	*HAN*	4	13.3109	0.0731
Forward	*SIC*	6	13.1292	0.0931
Forward	*P_corr*	8	13.9866	0.0118
GALMS	*AIC*	27	12.6532	0.1372
GALMS	*BIC*	20	13.6334	0.0771
GALMS	*SIC*	16	13.0012	0.0890

Tabella 13. Risultati ottenuti col modello lineare.

Il modello migliore sia in termini di scarto quadratico medio che di R^2 è quello ottenuto con il programma GALMS utilizzando come criterio selettivo l'AIC.

Un attento esame della distribuzione dei residui mostra una forte non linearità con evidenti effetti di eteroschedasticità.

Siccome i modelli statistici di durata sono espressi con funzioni di azzardo (hazard function)[8] si è utilizzato il seguente modello:

(28) $$h(t) = h_0(t)\exp(\sum_{i=1}^{n} a_i x_i)$$

dove $h(t)$ esprime la probabilità che si fuoriesca dalla situazione di disoccupazione nell'istante che va da t a $t + dt$, $h_o(t)$ la probabilità di transizione iniziale o funzione di rischio base (cioè la probabilità di transizione da uno stato di disoccupazione ad uno di occupazione indipendentemente dai valori assunti dalle variabili esplicative), $\exp(\sum_{i=1}^{n} a_i x_i)$ esprime l'impatto dei regressori selezionati su $h(t)$.

Si è scelta la forma log-lineare del modello a rischi proporzionali (modello di Cox) in quanto è una delle più utilizzate.

Indicando con y la durata della disoccupazione si può scrivere:

(29) $$y = e^{\sum_{i=1}^{n} a_i x_i} = \exp(\sum_{i=1}^{n} a_i x_i) \Rightarrow \ln y = \sum_{i=1}^{n} a_i x_i \,.$$

In questo modo si rende possibile il campionamento dei dati in tre gruppi (training set, validation set e test set) in quanto la variabilità della y viene ridotta sensibilmente (valori compresi tra –2.3 e +4.6) e con esso l'effetto delle "dummy" (rendendo possibile la stabilità del modello anche in caso di campionamento non bilanciato).

I risultati ottenuti sono riassunti in tabella 14. Anche in questo caso si è lavorato con i programmi "Backward", "Forward" e di selezione genetica come sopra.

---

[8] Se X è un numero aleatorio che rappresenta la durata di un certo fenomeno (per esempio il tempo di vita dell'oggetto di studio) e k una data temporale, la funzione di azzardo è definita come $h(k) = P(X = k \mid X \geq k)$. Tale funzione esprime la probabilità di morte dell'oggetto nel giorno k, data la sopravvivenza fino all'inizio del giorno k.

METODO	CRITERIO	DIMENSIONE	SQM	R^2
Backward	AIC	20	0.8932	0.1484
Backward	*BIC*	25	0.8881	0.1584
Backward	*HAN*	14	0.9162	0.1016
Backward	*SIC*	16	0.9034	0.1271
Backward	*p-value*	7	0.9213	0.0941
Forward	*HAN*	14	0.9162	0.1016
Forward	*SIC*	16	0.9137	0.1073
Forward	*P_corr*	5	1.1279	0.0880
GALMS	*SIC*	10	1.1773	0.1640

Tabella 14. Risultati ottenuti col modello esponenziale (linearizzato).

Confrontando i dati delle tabelle si nota come il modello esponenziale ottenuto dalla selezione genetica utilizzando come criterio selettivo l'indicatore SIC sia il migliore in assoluto in termini di R^2 (sebbene presenti uno scarto quadratico medio più alto).

Prima di trarre conclusioni sull'analisi svolta esaminiamo le variabili scelte per questo modello e la coerenza dei segni dei coefficienti calcolati.

Il modello seleziona 10 parametri (tutti significativi), che sono:

- L'età e l'età al quadrato degli individui intervistati, i cui coefficienti hanno segno opposto (rispettivamente 1.220 e – 0.162), compensandosi così vicendevolmente;
- Il numero dei componenti della famiglia di età inferiore ai 5 anni, che presenta un coefficiente negativo (- 0.219) facilmente spiegabile col fatto che in famiglie con bambini piccoli la ricerca di un lavoro cresce sensibilmente la necessità di mantenere i figli;
- Il numero dei componenti della famiglia, con coefficiente positivo (0.118) giustificato dal fatto che la durata della disoccupazione aumenta per le persone che vivono in famiglie numerose poiché vi è un fenomeno di compensazione con gli altri componenti;
- La condizione professionale finale, che indica se l'individuo fa parte del sottogruppo dei *disoccupati* in base alla definizione data in precedenza (appartengono alla forza lavoro). Il coefficiente negativo di questa variabile (- 0.243) è facilmente spiegabile in quanto un'individuo che non abbia mai smesso di cercare un lavoro è destinato a ricevere un'offerta più velocemente;
- L'uscita dal mercato del lavoro; questo parametro, che indica se l'individuo è stato messo fuori dal mercato del lavoro, mostra un coefficiente negativo (- 0.222) evidenziando che colui che è uscito da questo mercato può trovare un impiego più velocemente;
- L'uscita dal settore agricolo o edile, che indica se l'ultima attività è stata svolta nei due settori menzionati e presenta un coefficiente negativo (- 0.545) a dimostrazione che chi esce da questi settori si inserisce meglio nel mercato del lavoro;
- La preferenza del lavoro nel comune di residenza, che presenta un coefficiente positivo (0.304) il quale indica che quanto maggiori sono le richieste di questo tipo tanto più difficile è trovare un lavoro;
- Il percepimento di un sussidio, che sorprendentemente mostra un coefficiente negativo (- 0.361) evidenziando una certa concordanza con alcuni studi in questo settore [Nic79], [Gro90], [NaSt93]. D'altronde il sussidio potrebbe incentivare la non ricerca del lavoro fino a quando la scadenza del diritto a percepirlo è lontana; quando invece si avvicina, la ricerca potrebbe diventare più frenetica e l'accettazione dei posti di lavoro eventualmente offerti meno vincolata (coerentemente con la teoria del *job –search*).

- L'entrata nel settore agricolo o edile, con un coefficiente positivo (0.284) che dimostra che colui che ha in precedenza trovato un lavoro in questi settori presenta un periodo di disoccupazione maggiore.

L'analisi della funzione di ripartizione campionaria e teorica dei residui ottenuti utilizzando questo modello mostra un netto miglioramento del modello esponenziale rispetto a quello lineare puro.

Molto utili per una valutazione dei risultati sono anche i valori dei test di adattamento di Kolmogorov e di Cramér e von Mises dei tre modelli sopra studiati riportati in tabella 15.

Modello	**Criterio selettivo**	**Test di Kolmogorov**	**Test di Cramér e von Mises**
Lineare	AIC	0.1441	4.55
Esponenziale	*SIC*	0.0439	0.3135

Tabella 15. Risultati dei test di adattamento sui tre modelli esaminati (valori calcolati per alpha = 0.05).

Questi mostrano un progressivo miglioramento di adattamento alla distribuzione ipotizzata passando dal modello lineare a quello esponenziale.

Per il modello esponenziale è stato valutato anche l'indice di concordanza, *c*, considerando tutte le possibili coppie di osservazioni relative al test-set [Har84]. Per ogni data coppia la previsione si dice concordante se all'osservazione con una durata prevista maggiore corrisponde una durata osservata effettivamente più elevata. L'indice *c* corrisponde alla proporzione di concordanze su tutte le coppie possibili. Un indice $c=0.5$ rappresenta, quindi, un livello di concordanza associato ad un modello non predittivo, mentre un indice $c=1$ rappresenta un modello perfettamente predittivo (almeno dal punto di vista della concordanza). L'indice di concordanza per il modello esponenziale è risultato pari a 0.58.

6. CONCLUSIONI

In questo lavoro sono state presentate alcune possibili soluzioni alle problematiche inerenti alla selezione del modello statistico lineare multivariato orientato alla previsione.

Mentre la stima dei parametri associati al modello statistico prescelto avviene utilizzando gli approcci classici, per quanto concerne la determinazione del miglior set di regressori (quanti e quali) da fornire come input al modello si sono confrontati gli approcci classici con algoritmi evolutivi.

Si è visto che le tecniche classiche ("backward elimination" e "forward selection") soffrono di imposizioni strutturali (dipendenza dalla sequenza) e vincoli di carattere fisico la cui automazione li rende criticabili sotto più punti di vista.

Gli Algoritmi Genetici rimuovono le imposizioni strutturali (e con esse si eliminano i problemi connessi alla dipendenza dalla sequenza) valutando ogni soluzione rispetto alle proprie capacità di descrivere la soluzione del problema in esame sottoposta a un criterio di selezione e propagata alla "iterazione" successiva per mezzo di operatori che emulano la selezione genetica. Il processo evolutivo converge verso la migliore soluzione entro un criterio di stop preimposto.

Tutti gli algoritmi proposti sono stati testati su un particolare set di dati (valori di colesterolo ematico) che ha mostrato come l' approccio computazionale qui proposto risulti effettivamente più efficiente (non soltanto teoricamente). Sono stati valutati diversi criteri per la selezione del modello

e infine si è implementato un esempio semplice di sistema ibrido affiancando all'Algoritmo Genetico un criterio di scelta basato sulla Logica Fuzzy.

I risultati ottenuti, in ambito lineare, sono stati interessanti nonostante si sia sperimentato un algoritmo piuttosto semplice di cui si rende necessario un approfondimento per un'analisi più accurata.

L' approccio evolutivo e' stato utilizzato anche nell' ambito dell' analisi delle serie storiche. L'Algoritmo Genetico anche in questo caso ha risposto ottimamente alle aspettative. I modelli simulati sono stati correttamente specificati nella maggioranza dei casi e, quando questo non è avvenuto, è stato comunque scelto un modello che ha ottimizzato la procedura (*AIC* minore). Anche in questo caso sono stati valutati diversi criteri di selezione (AIC, BIC, SIC).

Se l'obiettivo finale era verificare il funzionamento degli Algoritmi Genetici e mettere a punto uno strumento che fosse in grado di prevedere correttamente anche in presenza di una componente di errore stocastica, i risultati sono sicuramente confortanti. Soltanto l'1% dei modelli simulati è stato scelto in modo errato senza la convergenza dell'algoritmo verso l'ottimo (minimo valore assunto dall'*AIC*). I risultati esposti nel paragrafo 4 mostrano che lo strumento utilizzato ha, complessivamente, buone capacità previsive.

Tutte le tipologie di algoritmi utilizzati sono stati testati su un largo set di dati simulati prima di essere applicati alla soluzione di problemi di natura finanziaria, ambientale e sociale.

In particolare, per quanto concerne questi ultimi due problemi, gli strumenti ottenuti sono stati applicati a due casi sperimentali riguardanti la previsione del livello di marea a Venezia del 1984 e il livello di disoccupazione registrato in Emilia Romagna tra il 1993 e il 1995.

I risultati forniti sia dalle simulazioni che dalle analisi sperimentali sono sicuramente interessanti sebbene spesso non è stato possibile approfondire le analisi per brevità o per i limiti computazionali imposti dalle macchine. I metodi utilizzati hanno comunque permesso di accertare alcune tendenze di tipo operativo: la sostituzione di persone esperte con tecnologie informatiche deve essere considerato solo in caso estremo, mentre lo scopo principale per cui si ricorre a tali tecniche è quello di incrementare l'efficienza e di migliorare i livelli di coerenza ed affidabilità dei processi decisionali sfruttando nel migliore dei modi i dati e le informazioni a disposizione; le procedure automatizzate con metodologie evolutive si sono rivelate più efficienti e robuste, nella generalità dei casi, rispetto a quelle classiche (anch'esse automatizzate)[9].

[9] Si ricorda infine che tutti gli strumenti implementati sono stati scritti in linguaggi MATLAB utilizzando, quando possibile, le funzioni inserite nei seguenti toolbox: Statistic toolbox, Genetic Algorithm toolbox e System Identification toolbox. Si è lavorato in ambiente Unix su macchine "Sun Ultrasparc2" (che permette una migliore approssimazione di calcolo rispetto ai PC con gli attuali sistemi operativi).

BIBLIOGRAFIA

[Akai69] Akaike A., *Statistical predictor identification*, Annals of the Institute of Statistical Mathematics 22, pp. 203-217, 1969.

[Akai73] Akaike A., *Information theory and an extension of the maximum likelihood principle.* In B.N. Petrov and F. Csai ed. "2nd International Symposium on Information Theory", pp. 267-281, Akademia Kiado, Budapest, 1973.

[Akai78] Akaike A., *A Bayesian analysis of the minimum AIC procedure*, Annals of the institute of Statistical Mathematics 30, part A, pp. 9-14, 1978.

[AV96] AAVV., *Using MATLAB*, The Math Works Inc., Natick, 1996.

[Bra97] Bradley J., *Statistic Toolbox User's Guide,* The Math Works Inc., Natick, 1997.

[BrLa94] Bremer R.H., Langevin G.J., *The Genetic Algorithm for identifying the structure of a fixed model*, 1994.

[CaPT86] Canestrelli P., Pastore F., Tommasin A., *Sviluppi di un modello operativo previsionale dellee maree di Venezia e revisione di casi rilevanti*, Comune di Venezia, Assessorato ai Trasporti e Servizi Pubblici, 1986.

[Chi95] Chipperfield A., Fleming P., Pohlheim H., Fonseca C., *Genetic Algorithm Toolbox User's Guide*, University of Sheffield, 1995.

[CLL96] Chaterjee S., Laudato M., Lynch L.A., *Genetic algorithms and their statistical applications: an introduction*, Computational Statistics & Data Analysis 22, pp. 633-651, 1996.

[Davi91] L.Davis,, *Handbook of Genetic Algorithm*, Van Nostrand Reinhold, New York, 1991.

[DrSm66] N.R.Draper, H. Smith, *Applied Regression Analysis, Wiley series in probability and mathematical statistics*, J. Wiley & Sons, New York, 1966.

[DuPr80] Dubois D., Prada H., *Fuzzy Sets and System: Theory and Application*, Academic Press, San Diego, 1980.

[GoKh95] Goonatilake S., Khebbal S., *Intelligent Hybrid System,* John Wiley, Chichester 1985.

[Gold89] Goldberg D.E., *Genetic Algorithms in search, optimization, and machine learning*, Addison-Wesley Publishing Corporation Inc., Reading (Mass.), 1989.

[Goli94] Golinelli R., *Metodi Econometrici di base per l'analisi delle serie storiche: alcune applicazioni pratiche sul personal computer*, Cluep, Bologna, 1994.

[Gre97] Greene W.H., *Econometric analysis*, McMillan, New York, 1997.

[Gro90] Groot W., *The Effects of Benefit and Duration Dependence on Re-Employment Probabilities*, Economic Letters, 32, 4, pp. 371-376, 1990.

[Hair95] Hair Joseph F. Jr., *Multivariate data analysis: with readings, Prentice-Hall*, Englewood Cliffs, 1995.

[HaQu79] Hannan E.J., Quinn B.G. *The determination of the order of an autoregression*, Journal of the Royal Statistical Society, B 41, pp. 190-195, 1979.

[Har84] Harrel et al., *Regression modelling strategies for improved prognostic predictions*, Statistics in Medicine, 3, pp. 143-152, 1984.

[Har93] Harvey A.C., *Time Series Model*, Harvester Wheatsheaf, New York, 1993.

[Holl75] Holland J. H., *Adaptation in Natural and Artificial Systems*, University of Michigan Press, AnnArbor, 1975.

[Holl95] J.H.Holland, *Hidden Order*, Reading, Addison Wesley Publishing Corporatin, Inc, 1995.

[Hod97] Hodges E.M., *Applications of Genetic Algorithms in Time Series Analysis,* 1997.

[HPS84] Hendry D.F., Pagan A.R., Sargan J.D., *Dynamic Specification*, in Griliches Z., Intriligator M.D. (a cura di), Handbook of Econometrics, vol. II, North Holland, 1984.

[HuTs89] Hurvich C.M., Tsai C.L., *Regression and time series model selection in small samples*, Biometrika 76, pp. 297-307, 1989.

[HuTS98] Hurvich C.M., Tsai C.L., Simonoff J.S., *Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion*, J of the Royal Society B (Statistical metodology) 60, pp. 271-293, 1998.

[LaPa99] Lalla M., Pattarin F., *Alcuni modelli per l'analisi delle durate complete e incomplete della disoccupazione: il caso Emilia-Romagna*, Working Paper n. 11, Dipartimento di Scienze Statistiche Università degli Studi di Padova, 1999.

[Lju95] Ljung L., *System Identification Toolbox User's Guide,* The Math Works Inc., Natick, 1995.

[Mal73] Mallows C.L., *Some comments on Cp*, Technometrics 15, pp. 661-675, 1973.

[MiPo97] T.Minerva, I.Poli, *Genetic Algorithms to identify Time Series Models",* Applied Stochastic Processes, Convegno di Capri, 1997.

[Mit96] M.Mitchell, *An Introduction to Genetic Algorithms*, Massachusetts, MIT Press,1996.

[NaSt93] Narendranathan W., Stewart M.B., *Modelling the Probability of Leaving Unemployment: competing Risks Models with Flexible Base-line Hazards,* J. of the Royal Statistical Society C (Applied Statistics), 42, pp. 63-83, 1993.

[Nic79] Nickel S., *The Effect of Unemployment and Related Benefits on the Duration of Unemployment*, Economic Journal 89, pp. 63-83, 1979.

[Orsi95] Orsi R., *Probabilità e inferenza statistica*, Il Mulino, Bologna, 1995.

[Picc74] Piccolo D., *Analisi delle serie temporali: i processi autoregressivi del secondo ordine*, Centro di specializzazione e ricerche economico-agrarie per il mezzogiorno, Portici, 1974.

[Picc90] Piccolo D., *Introduzione all'analisi delle serie storiche*, La Nuova Scientifica, Roma, 1990.

[Picc98] Piccolo D., *Statistica*, Il Mulino, Bologna, 1998.

[Prie81] Priestley M.B., *Spectral Analysis and Time Series*, Vols. 1 e 2, Academic Press, New York, 1981.

[QuTs98] McQuarrie Allan D.R., Tsai Chih-Ling, *Regression and Time Series Model Selection*, World Scientific, Singapore, 1998.

[Rissa78] Rissanen J., *Modeling by Shortest Data Description,* Automatica 14, pp. 465-471, 1978.

[Schw78] Schwarz G., *Estimating the dimension of a model*, Annals of Statistics 6, pp. 461-464, 1978.

[SeSr90] Sen A., Srivastava M., *Regression Analysis. Theory, Methods and Aplications*, Springer-Verlag, New York, 1990.

[ShTs98] Shi P., Tsai Chih-Ling, *A note on the unification of the Akaike information criterion*, J. of the Royal Statistical Society B (Statistical Metodology) 60, pp. 551-558, 1998.

[The61] Theil H, *Economic Foprecasts and Policy*, North-Holland Publ. Co., Amsterdam, 1961.

[The67] Theil H, *Economics and Information Theory*, North-Holland Publ. Co., Amsterdam, 1967.

[Whi92] White D., Sofge D., *Handbook of Intelligent Control*, New York, Van Nostrand Reinhold, 1992.

[Zade65] Zadeh L.A., *Fuzzy Sets*, Information and Control, Vol.8, pp. 338-353, 1965.

[Zade84] Zadeh L.A., *Making Computers Think Like People,* IEEE Spectrum, Vol. 21:8, pp. 26-32, 1984.

[Zade94] Zadeh L.A., *Fuzzy Logic and Soft Computing: Issues, Contentions and Perspectives*, In:,IIZUKA'94:3rd International Conference on Fuzzy Logic, Neural Nets and Soft Computing, pp. 1-2, Iisuka, Japan 1994.

1. Maria Cristina Marcuzzo [1985] "Yoan Violet Robinson (1903-1983)", pp. 134

2. Sergio Lugaresi [1986] "Le imposte nelle teorie del sovrappiù", pp. 26

3. Massimo D'Angelillo e Leonardo Paggi [1986] "PCI e socialdemocrazie europee. Quale riformismo?", pp. 158

4. Gian Paolo Caselli e Gabriele Pastrello [1986] "Un suggerimento hobsoniano su terziario ed occupazione: il caso degli Stati Uniti 1960/1983", pp. 52

5. Paolo Bosi e Paolo Silvestri [1986] "La distribuzione per aree disciplinari dei fondi destinati ai Dipartimenti, Istituti e Centri dell'Università di Modena: una proposta di riforma", pp. 25

6. Marco Lippi [1986] "Aggregations and Dynamic in One-Equation Econometric Models", pp. 64

7. Paolo Silvestri [1986] "Le tasse scolastiche e universitarie nella Legge Finanziaria 1986", pp. 41

8. Mario Forni [1986] "Storie familiari e storie di proprietà. Itinerari sociali nell'agricoltura italiana del dopoguerra", pp. 165

9. Sergio Paba [1986] "Gruppi strategici e concentrazione nell'industria europea degli elettrodomestici bianchi", pp. 56

10. Nerio Naldi [1986] "L'efficienza marginale del capitale nel breve periodo", pp. 54

11. Fernando Vianello [1986] "Labour Theory of Value", pp. 31

12. Piero Ganugi [1986] "Risparmio forzato e politica monetaria negli economisti italiani tra le due guerre", pp. 40

13. Maria Cristina Marcuzzo e Annalisa Rosselli [1986] "The Theory of the Gold Standard and Ricardo's Standard Comodity", pp. 30

14. Giovanni Solinas [1986] "Mercati del lavoro locali e carriere di lavoro giovanili", pp. 66

15. Giovanni Bonifati [1986] "Saggio dell'interesse e domanda effettiva. Osservazioni sul cap. 17 della General Theory", pp. 42

16. Marina Murat [1986] "Betwin old and new classical macroeconomics: notes on Lejonhufvud's notion of full information equilibrium", pp. 20

17. Sebastiano Brusco e Giovanni Solinas [1986] "Mobilità occupazionale e disoccupazione in Emilia Romagna", pp. 48

18. Mario Forni [1986] "Aggregazione ed esogeneità", pp. 13

19. Sergio Lugaresi [1987] "Redistribuzione del reddito, consumi e occupazione", pp. 17

20. Fiorenzo Sperotto [1987] "L'immagine neopopulista di mercato debole nel primo dibattito sovietico sulla pianificazione", pp. 34

21. M. Cecilia Guerra [1987] "Benefici tributari nel regime misto per i dividendi proposto dalla commissione Sarcinelli: una nota critica", pp. 9

22. Leonardo Paggi [1987] "Contemporary Europe and Modern America: Theories of Modernity in Comparative Perspective", pp. 38

23. Fernando Vianello [1987] "A Critique of Professor Goodwin's 'Critique of Sraffa'", pp. 12

24. Fernando Vianello [1987] "Effective Demand and the Rate of Profits. Some Thoughts on Marx, Kalecki and Sraffa", pp. 41

25. Anna Maria Sala [1987] "Banche e territorio. Approccio ad un tema geografico-economico", pp. 40

26. Enzo Mingione e Giovanni Mottura [1987] "Fattori di trasformazione e nuovi profili sociali nell'agricoltura italiana: qualche elemento di discussione", pp. 36

27. Giovanna Procacci [1988] "The State and Social Control in Italy During the First World War", pp. 18

28. Massimo Matteuzzi e Annamaria Simonazzi [1988] "Il debito pubblico", pp. 62

29. Maria Cristina Marcuzzo (a cura di) [1988] "Richard F. Kahn. A discipline of Keynes", pp. 118

30. Paolo Bosi [1988] "MICROMOD. Un modello dell'economia italiana per la didattica della politica fiscale", pp. 34

31. Paolo Bosi [1988] "Indicatori della politica fiscale. Una rassegna e un confronto con l'aiuto di MICROMOD", pp. 25

32. Giovanna Procacci [1988] "Protesta popolare e agitazioni operaie in Italia 1915-1918", pp. 45

33. Margherita Russo [1988] "Distretto Industriale e servizi. Uno studio dei trasporti nella produzione e nella vendita delle piastrelle", pp. 157

34. Margherita Russo [1988] "The effect of technical change on skill requirements: an empirical analysis", pp. 28

35. Carlo Grillenzoni [1988] "Identification, estimations of multivariate transfer functions", pp. 33

36. Nerio Naldi [1988] "'Keynes' concept of capital", pp. 40

37. Andrea Ginzburg [1988] "locomotiva Italia?", pp. 30

38. Giovanni Mottura [1988] "La 'persistenza' secolare. Appunti su agricoltura contadina ed agricoltura familiare nelle società industriali", pp. 40

39. Giovanni Mottura [1988] "L'anticamera dell'esodo. I contadini italiani della 'restaurazione contrattuale' fascista alla riforma fondiaria", pp. 40

40. Leonardo Paggi [1988] "Americanismo e riformismo. La socialdemocrazia europea nell'economia mondiale aperta", pp. 120

41. Annamaria Simonazzi [1988] "Fenomeni di isteresi nella spiegazione degli alti tassi di interesse reale", pp. 44

42. Antonietta Bassetti [1989] "Analisi dell'andamento e della casualità della borsa valori", pp. 12

43. Giovanna Procacci [1989] "State coercion and worker solidarity in Italy (1915-1918): the moral and political content of social unrest", pp. 41

44. Carlo Alberto Magni [1989] "Reputazione e credibilità di una minaccia in un gioco bargaining", pp. 56

45. Giovanni Mottura [1989] "Agricoltura familiare e sistema agroalimentare in Italia", pp. 84

46. Mario Forni [1989] "Trend, Cycle and 'Fortuitous cancellation': a Note on a Paper by Nelson and Plosser", pp. 4

47. Paolo Bosi , Roberto Golinelli , Anna Stagni [1989] "Le origini del debito pubblico e il costo della stabilizzazione", pp. 26

48. Roberto Golinelli [1989] "Note sulla struttura e sull'impiego dei modelli macroeconometrici", pp. 21

49. Marco Lippi [1989] "A Shorte Note on Cointegration and Aggregation", pp. 11

50. Gian Paolo Caselli e Gabriele Pastrello [1989] "The Linkage between Tertiary and Industrial Sector in the Italian Economy: 1951-1988. From an External Dependence to an International One", pp. 40

51. Gabriele Pastrello [1989] "Francois quesnay: dal Tableau Zig-zag al Tableau Formule: una ricostruzione", pp. 48

52. Paolo Silvestri [1989] "Il bilancio dello stato", pp. 34

53. Tim Mason [1990] "Tre seminari di storia sociale contemporanea", pp. 26

54. Michele Lalla [1990] "The Aggregate Escape Rate Analysed throught the Queueing Model", pp. 23

55. Paolo Silvestri [1990] "Sull'autonomia finanziaria dell'università", pp. 11

56. Paola Bertolini, Enrico Giovannetti [1990] "Uno studio di 'filiera' nell'agroindustria. Il caso del Parmigiano Reggiano", pp. 164

57. Paolo Bosi, Roberto Golinelli, Anna Stagni [1990] "Effetti macroeconomici, settoriali e distributivi dell'armonizzazione dell'IVA", pp. 24

58. Michele Lalla [1990] "Modelling Employment Spells from Emilia Labour Force Data", pp. 18

59. Andrea Ginzburg [1990] "Politica Nazionale e commercio internazionale", pp. 22

60. Andrea Giommi [1990] "La probabilità individuale di risposta nel trattamento dei dati mancanti", pp. 13

61. Gian Paolo Caselli e Gabriele Pastrello [1990] "The service sector in planned economies. Past experiences and future prospectives", pp. 32

62. Giovanni Solinas [1990] "Competenze, grandi industrie e distretti industriali,. Il caso Magneti Marelli", pp. 23

63. Andrea Ginzburg [1990] "Debito pubblico, teorie monetarie e tradizione civica nell'Inghilterra del Settecento", pp. 30

64. Mario Forni [1990] "Incertezza, informazione e mercati assicurativi: una rassegna", pp. 37

65. Mario Forni [1990] "Misspecification in Dynamic Models", pp. 19

66. Gian Paolo Caselli e Gabriele Pastrello [1990] "Service Sector Growth in CPE's: An Unsolved Dilemma", pp. 28

67. Paola Bertolini [1990] "La situazione agro-alimentare nei paesi ad economia avanzata", pp. 20

68. Paola Bertolini [1990] "Sistema agro-alimentare in Emilia Romagna ed occupazione", pp. 65

69. Enrico Giovannetti [1990] "Efficienza ed innovazione: il modello "fondi e flussi" applicato ad una filiera agro-industriale", pp. 38

70. Margherita Russo [1990] "Cambiamento tecnico e distretto industriale: una verifica empirica", pp. 115

71. Margherita Russo [1990] "Distretti industriali in teoria e in pratica: una raccolta di saggi", pp. 119

72. Paolo Silvestri [1990] " La Legge Finanziaria. Voce dell'enciclopedia Europea Garzanti", pp. 8

73. Rita Paltrinieri [1990] "La popolazione italiana: problemi di oggi e di domani", pp. 57

74. Enrico Giovannetti [1990] "Illusioni ottiche negli andamenti delle Grandezze distributive: la scala mobile e l''appiattimento' delle retribuzioni in una ricerca", pp. 120

75. Enrico Giovannetti [1990] "Crisi e mercato del lavoro in un distretto industriale: il bacino delle ceramiche. Sez I", pp. 150

76. Enrico Giovannetti [1990] " Crisi e mercato del lavoro in un distretto industriale: il bacino delle ceramiche. Sez. II", pp. 145

78. Antonietta Bassetti e Costanza Torricelli [1990] "Una riqualificazione dell'approccio bargaining alla selezioni di portafoglio", pp. 4

77. Antonietta Bassetti e Costanza Torricelli [1990] "Il portafoglio ottimo come soluzione di un gioco bargaining", pp. 15

79. Mario Forni [1990] "Una nota sull'errore di aggregazione", pp. 6

80. Francesca Bergamini [1991] "Alcune considerazioni sulle soluzioni di un gioco bargaining", pp. 21

81. Michele Grillo e Michele Polo [1991] "Political Exchange and the allocation of surplus: a Model of Two-party competition", pp. 34

82. Gian Paolo Caselli e Gabriele Pastrello [1991] "The 1990 Polish Recession: a Case of Truncated Multiplier Process", pp. 26

83. Gian Paolo Caselli e Gabriele Pastrello [1991] "Polish firms: Pricate Vices Pubblis Virtues", pp. 20

84. Sebastiano Brusco e Sergio Paba [1991] "Connessioni, competenze e capacità concorrenziale nell'industria della Sardegna", pp. 25

85. Claudio Grimaldi, Rony Hamaui, Nicola Rossi [1991] "Non Marketable assets and hauseholds' Portfolio Choice: a Case of Study of Italy", pp. 38

86. Giulio Righi, Massimo Baldini, Alessandra Brambilla [1991] "Le misure degli effetti redistributivi delle imposte indirette: confronto tra modelli alternativi", pp. 47

87. Roberto Fanfani, Luca Lanini [1991] "Innovazione e servizi nello sviluppo della meccanizzazione agricola in Italia", pp. 35

88. Antonella Caiumi e Roberto Golinelli [1992] "Stima e applicazioni di un sistema di domanda Almost Ideal per l'economia italiana", pp. 34

89. Maria Cristina Marcuzzo [1992] "La relazione salari-occupazione tra rigidità reali e rigidità nominali", pp. 30

90. Mario Biagioli [1992] "Employee financial participation in enterprise results in Italy", pp. 50

91. Mario Biagioli [1992] "Wage structure, relative prices and international competitiveness", pp. 50

92. Paolo Silvestri e Giovanni Solinas [1993] "Abbandoni, esiti e carriera scolastica. Uno studio sugli studenti iscritti alla Facoltà di Economia e Commercio dell'Università di Modena nell'anno accademico 1990/1991", pp. 30

93. Gian Paolo Caselli e Luca Martinelli [1993] "Italian GPN growth 1890-1992: a unit root or segmented trend representatin?", pp. 30

94. Angela Politi [1993] "La rivoluzione fraintesa. I partigiani emiliani tra liberazione e guerra fredda, 1945-1955", pp. 55

95. Alberto Rinaldi [1993] "Lo sviluppo dell'industria metalmeccanica in provincia di Modena: 1945-1990", pp. 70

96. Paolo Emilio Mistrulli [1993] "Debito pubblico, intermediari finanziari e tassi d'interesse: il caso italiano", pp. 30

97. Barbara Pistoresi [1993] "Modelling disaggregate and aggregate labour demand equations. Cointegration analysis of a labour demand function for the Main Sectors of the Italian Economy: 1950-1990", pp. 45

98. Giovanni Bonifati [1993] "Progresso tecnico e accumulazione di conoscenza nella teoria neoclassica della crescita endogena. Una analisi critica del modello di Romer", pp. 50

99. Marcello D'Amato e Barbara Pistoresi [1994] "The relationship(s) among Wages, Prices, Unemployment and Productivity in Italy", pp. 30

100. Mario Forni [1994] "Consumption Volatility and Income Persistence in the Permanent Income Model", pp. 30

101. Barbara Pistoresi [1994] "Using a VECM to characterise the relative importance of permanent and transitority components", pp. 28

102. Gian Paolo Caselli and Gabriele Pastrello [1994] "Polish recovery form the slump to an old dilemma", pp. 20

103. Sergio Paba [1994] "Imprese visibili, accesso al mercato e organizzazione della produzione", pp. 20

104. Giovanni Bonifati [1994] "Progresso tecnico, investimenti e capacità produttiva", pp. 30

105. Giuseppe Marotta [1994] "Credit view and trade credit: evidence from Italy", pp. 20

106. Margherita Russo [1994] "Unit of investigation for local economic development policies", pp. 25

107. Luigi Brighi [1995] "Monotonicity and the demand theory of the weak axioms", pp. 20

108. Mario Forni e Lucrezia Reichlin [1995] "Modelling the impact of technological change across sectors and over time in manufactoring", pp. 25

109. Marcello D'Amato and Barbara Pistoresi [1995] "Modelling wage growth dynamics in Italy: 1960-1990", pp. 38

110. Massimo Baldini [1995] "INDIMOD. Un modello di microsimulazione per lo studio delle imposte indirette", pp. 37

111. Paolo Bosi [1995] "Regionalismo fiscale e autonomia tributaria: l'emersione di un modello di consenso", pp. 38

112. Massimo Baldini [1995] "Aggregation Factors and Aggregation Bias in Consumer Demand", pp. 33

113. Costanza Torricelli [1995] "The information in the term structure of interest rates. Can stocastic models help in resolving the puzzle?" pp. 25

114. Margherita Russo [1995] "Industrial complex, pôle de développement, distretto industriale. Alcune questioni sulle unità di indagine nell'analisi dello sviluppo." pp. 45

115. Angelika Moryson [1995] "50 Jahre Deutschland. 1945 - 1995" pp. 21

116. Paolo Bosi [1995] "Un punto di vista macroeconomico sulle caratteristiche di lungo periodo del nuovo sistema pensionistico italiano." pp. 32

117. Gian Paolo Caselli e Salvatore Curatolo [1995] "Esistono relazioni stimabili fra dimensione ed efficienza delle istituzioni e crescita produttiva? Un esercizio nello spirito di D.C. North." pp. 11

118. Mario Forni e Marco Lippi [1995] "Permanent income, heterogeneity and the error correction mechanism." pp. 21

119. Barbara Pistoresi [1995] "Co-movements and convergence in international output. A Dynamic Principal Components Analysis" pp. 14

120. Mario Forni e Lucrezia Reichlin [1995] "Dynamic common factors in large cross-section" pp. 17

121. Giuseppe Marotta [1995] "Il credito commerciale in Italia: una nota su alcuni aspetti strutturali e sulle implicazioni di politica monetaria" pp. 20

122. Giovanni Bonifati [1995] "Progresso tecnico, concorrenza e decisioni di investimento: una analisi delle determinanti di lungo periodo degli investimenti" pp. 25

123. Giovanni Bonifati [1995] "Cambiamento tecnico e crescita endogena: una valutazione critica delle ipotesi del modello di Romer" pp. 21

124. Barbara Pistoresi e Marcello D'Amato [1995] "La riservatezza del banchiere centrale è un bene o un male? ‚Effetti dell'informazione incompleta sul benessere in un modello di politica monetaria." pp. 32

125. Barbara Pistoresi [1995] "Radici unitarie e persistenza: l'analisi univariata delle fluttuazioni economiche." pp. 33

126. Barbara Pistoresi e Marcello D'Amato [1995] "Co-movements in European real outputs" pp. 20

127. Antonio Ribba [1996] "Ciclo economico, modello lineare-stocastico, forma dello spettro delle variabili macroeconomiche" pp. 31

128. Carlo Alberto Magni [1996] "Repeatable and una tantum real options a dynamic programming approach" pp. 23

129. Carlo Alberto Magni [1996] "Opzioni reali d'investimento e interazione competitiva: programmazione dinamica stocastica in optimal stopping" pp. 26

130. Carlo Alberto Magni [1996] "Vaghezza e logica fuzzy nella valutazione di un'opzione reale" pp. 20

131. Giuseppe Marotta [1996] "Does trade credit redistribution thwart monetary policy? Evidence from Italy" pp. 20

132. Mauro Dell'Amico e Marco Trubian [1996] "Almost-optimal solution of large weighted equicut problems" pp. 30

133. Carlo Alberto Magni [1996] "Un esempio di investimento industriale con interazione competitiva e avversione al rischio" pp. 20

134. Margherita Russo, Peter Börkey, Emilio Cubel, François Lévêque, Francisco Mas [1996] "Local sustainability and competitiveness: the case of the ceramic tile industry" pp. 66

135. Margherita Russo [1996] "Camionetto tecnico e relazioni tra imprese" pp. 190

136. David Avra Lane, Irene Poli, Michele Lalla, Alberto Roverato [1996] "Lezioni di probabilità e inferenza statistica" pp. 288

137. David Avra Lane, Irene Poli, Michele Lalla, Alberto Roverato [1996] "Lezioni di probabilità e inferenza statistica - Esercizi svolti - " pp. 302

138. Barbara Pistoresi [1996] "Is an Aggregate Error Correction Model Representative of Disaggregate Behaviours? An example" pp. 24

139. Luisa Malaguti e Costanza Torricelli [1996] "Monetary policy and the term structure of interest rates" , pp. 30

140. Mauro Dell'Amico, Martine Labbé, Francesco Maffioli [1996] "Exact solution of the SONET Ring Loading Problem", pp. 20

141. Mauro Dell'Amico, R.J.M. Vaessens [1996] "Flow and open shop scheduling on two machines with transportation times and machine-independent processing times in NP-hard, pp. 10

142. M. Dell'Amico, F. Maffioli, A. Sciomechen [1996] "A Lagrangean Heuristic for the Pirze Collecting Travelling Salesman Problem", pp. 14

143. Massimo Baldini [1996] "Inequality Decomposition by Income Source in Italy - 1987 - 1993", pp. 20

144. Graziella Bertocchi [1996] "Trade, Wages, and the Persistence of Underdevelopment" pp. 20

145. Graziella Bertocchi and Fabio Canova [1996] "Did Colonization matter for Growth? An Empirical Exploration into the Historical Causes of Africa's Underdevelopment" pp. 32

146. Paola Bertolini [1996] "La modernization de l'agricolture italienne et le cas de l'Emilie Romagne" pp. 20

147. Enrico Giovannetti [1996] "Organisation industrielle et développement local: le cas de l'agroindutrie in Emilie Romagne" pp. 18

148. Maria Elena Bontempi e Roberto Golinelli [1996] "Le determinanti del leverage delle imprese: una applicazione empirica ai settori industriali dell'economia italiana" pp. 31

149. Paola Bertolini [1996] "L'agriculture et la politique agricole italienne face aux recents scenarios", pp. 20

150. Enrico Giovannetti [1996] "Il grado di utilizzo della capacità produttiva come misura dei costi di transazione: una rilettura di 'Nature of the Firm' di R. Coase", pp. 75

151. Enrico Giovannetti [1996] "Il I° ciclo del Diploma Universitario Economia e Amministrazione delle Imprese", pp. 25

152. Paola Bertolini, Enrico Giovannetti, Giulia Santacaterina [1996] "Il Settore del Verde Pubblico. Analisi della domanda e valutazione economica dei benefici", pp. 35

153. Giovanni Solinas [1996] "Sistemi produttivi del Centro-Nord e del Mezzogiorno. L'industria delle calzature", pp. 55

154. Tindara Addabbo [1996] "Married Women's Labour Supply in Italy in a Regional Perspective", pp. 85

155. Paolo Silvestri, Giuseppe Catalano, Cristina Bevilacqua [1996] "Le tasse universitarie e gli interventi per il diritto allo studio: la prima fase di applicazione di una nuova normativa" pp. 159

156. Sebastiano Brusco, Paolo Bertossi, Margherita Russo [1996] "L'industria dei rifiuti urbani in Italia", pp. 25

157. Paolo Silvestri, Giuseppe Catalano [1996] "Le risorse del sistema universitario italiano: finanziamento e governo" pp. 400

158. Carlo Alberto Magni [1996] "Un semplice modello di opzione di differimento e di vendita in ambito discreto", pp. 10

159. Tito Pietra, Paolo Siconolfi [1996] "Fully Revealing Equilibria in Sequential Economies with Asset Markets" pp. 17

160. Tito Pietra, Paolo Siconolfi [1996] "Extrinsic Uncertainty and the Informational Role of Prices" pp. 42

161. Paolo Bertella Farnetti [1996] "Il negro e il rosso. Un precedente non esplorato dell'integrazione afroamericana negli Stati Uniti" pp. 26

162. David Lane [1996] "Is what is good for each best for all? Learning from others in the information contagion model" pp. 18

163. Antonio Ribba [1996] "A note on the equivalence of long-run and short-run identifying restrictions in cointegrated systems" pp. 10

164. Antonio Ribba [1996] "Scomposizioni permanenti-transitorie in sistemi cointegrati con una applicazione a dati italiani" pp. 23

165. Mario Forni, Sergio Paba [1996] "Economic Growth, Social Cohesion and Crime" pp. 20

166. Mario Forni, Lucrezia Reichlin [1996] "Let's get real: a factor analytical approch to disaggregated business cycle dynamics" pp. 25

167. Marcello D'Amato e Barbara Pistoresi [1996] "So many Italies: Statistical Evidence on Regional Cohesion" pp. 31

168. Elena Bonfiglioli, Paolo Bosi, Stefano Toso [1996] "L'equità del contributo straordinario per l'Europa" pp. 20

169. Graziella Bertocchi, Michael Spagat [1996] "Il ruolo dei licei e delle scuole tecnico-professionali tra progresso tecnologico, conflitto sociale e sviluppo economico" pp. 37

170. Gianna Boero, Costanza Torricelli [1997] "The Expectations Hypothesis of the Term Structure of Interest Rates: Evidence for Germany" pp. 15

171. Mario Forni, Lucrezia Reichlin [1997] "National Policies and Local Economies: Europe and the US" pp. 22

172. Carlo Alberto Magni [1997] "La trappola del Roe e la tridimensionalità del Van in un approccio sistemico", pp. 16

173. Mauro Dell'Amico [1997] "A Linear Time Algorithm for Scheduling Outforests with Communication Delays on Two or Three Processor"pp. 18

174. Paolo Bosi [1997] "Aumentare l'età pensionabile fa diminuire la spesa pensionistica? Ancora sulle caratteristiche di lungo periodo della riforma Dini" pp. 13

175. Paolo Bosi e Massimo Matteuzzi [1997] "Nuovi strumenti per l'assistenza sociale" pp 31

176. Mauro Dell'Amico, Francesco Maffioli e Marco Trubian [1997] "New bounds for optium traffic assignment in satellite communication" pp. 21

177. Carlo Alberto Magni [1997] "Paradossi, inverosimiglianze e contraddizioni del Van: operazioni certe" pp. 9

178. Barbara Pistoresi e Marcello D'Amato [1997] "Persistence of relative unemployment rates across italian regions" pp. 25

179. Margherita Russo, Franco Cavedoni e Riccardo Pianesani [1997] "Le spese ambientali dei Comuni in provincia di Modena, 1993-1995" pp. 23

180. Gabriele Pastrello [1997] "Time and Equilibrium, Two Elisive Guests in the Keynes-Hawtrey-Robertson Debate in the Thirties" pp. 25

181. Luisa Malaguti e Costanza Torricelli [1997] "The Interaction Between Monetary Policy and the Expectation Hypothesis of the Term Structure of Interest rates in a N-Period Rational Expectation Model" pp. 27

182. Mauro Dell'Amico [1997] "On the Continuous Relaxation of Packing Problems – Technical Note" pp. 8

183. Stefano Bordoni [1997] "Prova di Idoneità di Informatica Dispensa Esercizi Excel 5" pp 49

184. Francesca Bergamini e Stefano Bordoni [1997] "Una verifica empirica di un nuovo metodo di selezione ottima di portafoglio" pp. 22

185. Gian Paolo Caselli e Maurizio Battini [1997] "Following the tracks of atkinson and micklewright the changing distribution of income and earnings in poland from 1989 to 1995".pp 21

186. Mauro Dell'Amico e Francesco Maffioli [1997] "Combining Linear and Non-Linear Objectives in Spanning Tree Problems" pp. 21

187. Gianni Ricci e Vanessa Debbia [1997] "Una soluzione evolutiva in un gioco differenziale di lotta di classe" pp.14

188. Fabio Canova e Eva Ortega [1997] "Testing Calibrated General Equilibrium Model" pp 34

189. Fabio Canova [1997] "Does Detrending Matter for the Determination of the Reference Cycle and the Selection of Turning Points?" pp. 35

190. Fabio Canova e Gianni De Nicolò [1997] "The Equity Premium and the Risk Free Rate: A Cross Country, Cross Maturity Examination" pp. 41

191. Fabio Canova e Angel J. Ubide [1997] "International Business Cycles, Financial Market and Household Production" pp. 32

192. Fabio Canova e Gianni De Nicolò [1997] "Stock Returns, Term Structure, Inflation and Real Activity: An International Perspective" pp. 33

193. Fabio Canova e Morten Ravn [1997] "The Macroeconomic Effects of German Unification: Real Adjustmentsand the Welfare State" pp. 34

194. Fabio Canova [1997] "Detrending and Business Cycle Facts" pp. 40

195. Fabio Canova e Morten O. Ravn [1997] "Crossing the Rio Grande: Migrations, Business Cycle and the Welfare State" pp. 37

196. Fabio Canova e Jane Marrinan [1997] "Sources and Propagation of International Output Cycles: Common Shocks or Transmission?" pp. 41

197. Fabio Canova e Albert Marcet [1997] "The Poor Stay Poor: Non-Convergence Across Countries and Regions" pp. 44

198. Carlo Alberto Magni [1997] "Un Criterio Strutturalista per la Valutazione di Investimenti" pp. 17

199. Stefano Bordoni [1997] "Elaborazione Automatica dei Dati" pp. 60

200. Paolo Bertella Farnetti [1997] "The United States and the Origins of European Integration" pp. 19

201. Paolo Bosi [1997] "Sul Controllo Dinamico di un Sistema Pensionistico a Ripartizione di Tipo Contributivo" pp 17

202. Paola Bertolini [1997] "European Union Agricultural Policy: Problems and Perspectives" pp18

203. Stefano Bordoni [1997] "Supporti Informatici per la Ricerca delle soluzioni di Problemi Decisionali" pp30

204. Carlo Alberto Magni [1997] "Paradossi, Inverosimiglianze e Contraddizioni del Van: Operazioni Aleatorie" pp10

205. Carlo Alberto Magni [1997] "Tir, Roe e Van: Distorsioni linguistiche e Cognitive nella Valutazione degli Investimenti" pp 17

206. Gisella Facchinetti, Roberto Ghiselli Ricci e Silvia Muzzioli [1997] "New Methods For Ranking Triangular Fuzzy Numbers: An Investment Choice" pp 9

207. Mauro Dell'Amico e Silvano Martello [1997] "Reduction of the Three-Partition Problem" pp16

208. Carlo Alberto Magni [1997] "IRR, ROE and NPV: a Systemic Approach" pp. 20

209. Mauro Dell'Amico, Andrea Lodi e Francesco Maffioli [1997] "Solution of the cumulative assignment problem with a well-structured tabu search method" pp. 25

210. Carlo Alberto Magni [1997] "La definizione di investimento e criterio del Tir ovvero: la realtà inventata" pp.16

211. Carlo Albero Magni [1997] "Critica alla definizione classica di investimento: un approccio sistemico" pp17

212. Alberto Roverato [1997] "Asymptotic prior to posterior analysis for graphical gaussian models" pp.8

213. Tindara Addabbo [1997] "Povertà nel 1995 analisi statica e dinamica sui redditi familiari" pp 64

214. Gian Paolo Caselli e Franca Manghi [1997] "La transizione da piano a mercato e il modello di Ising" pp15

215. Tindara Addabbo [1998] "Lavoro non pagato e reddito esteso: un'applicazione alle famiglie italiane in cui entrambi i coniugi sono lavoratori dipendenti" pp 54

216. Tindara Addabbo [1998] "Probabilità di occupazione e aspettative individuali" pp 36

217. Lara Magnani [1998] "Transazioni, contratti e organizzazioni: una chiave di lettura della teoria economica dell'organizzazione pp 39

218. Michele Lalla, Rosella Molinari e Maria Grazia Modena [1998] "La progressione delle carriere: i percorsi in cardiologia" pp 46

219. Lara Magnani [1998] "L'organizzazione delle transazioni di subfornitura nel distretto industriale" pp 40

220. Antonio Ribba [1998] "Recursive VAR orderings and identification of permanent and transitory shocks" pp12

221. Antonio Ribba [1998] "Granger-causality and exogeneity in cointegrated Var models" pp 5

222. Luigi Brighi e Marcello D'Amato [1998] "Optimal Procurement in Multiproduct Monopoly" pp 25

223. Paolo Bosi, Maria Cecilia Guerra e Paolo Silvestri [1998] "La spesa sociale nel comune Modena" Rapporto intermedio pp 37

224. Mario Forni e Marco Lippi [1998] "On the Microfoundations of Dynamic Macroeconomics" pp 22

225. Roberto Ghiselli Ricci [1998] "Nuove Proposte di Ordinamento di Numeri Fuzzy.Una Applicazione ad un Problema di Finanziamento pp 7

226. Tommaso Minerva [1998] "Internet Domande e Risposte" pp 183

227. Tommaso Minerva [1998] "Elementi di Statistica Computazione. Parte Prima: .Il Sistema Operativo Unix ed il Linguaggio C" pp. 57

228. Tommaso Minerva and Irene Poli [1998] "A Gennetic Algorithms Selection Method for Predictive Neural Nets and Linear Modenls" pp. 60

229. Tommaso Minerva and Irene Poli [1998] "Building an ARMA Model by using a Genetic Algorithm" pp. 60

230. Mauro Dell'Amico e Paolo Toth [1998] "Algorithms and Codes for Dense Assignment Problems: the State of the Art" pp 35

231. Ennio Cavazzuti e Nicoletta Pacchiarotti [1998] "How to play an hotelling game in a square town" pp 12

232. Alberto Roverato e Irene Poli [1998] "Un algoritmo genetico per la selezione di modelli grafici" pp 11

233. Marcello D'Amato e Barbara Pistoresi [1998] "Delegation of Monetary Policy to a Central Banker with Private Information" pp 15.

234. Graziella Bertocchi e Michael Spagat [1998] "The Evolution of Modern Educational Systems. Technical vs. General Education, Distributional Conflict, and Growth" pp 31

235. Andrè Dumas [1998] "Le systeme monetaire Europeen" pp 24.

236. Gianna Boero, Gianluca Di Lorenzo e Costanza Torricelli [1998] "The influence of short rate predictability and monetary policy on tests of the expectations hypothesis: some comparative evidence" pp 30

237. Carlo Alberto Magni [1998] "A systemic rule for investment decisions: generalizations of the traditional DCF criteria and new conceptions" pp 30

238. Marcello D'Amato e Barbara Pistoresi [1998] "Interest Rate Spreads Between Italy and Germany: 1995-1997" pp 16

239. Paola Bertolini e Alberto Bertacchini [1998] "Il distretto di lavorazioni carni suine in provincia di Modena" pp 29

240. Costanza Torricelli e Gianluca Di Lorenzo [1998] "Una nota sui fondamenti matematico-finanziari della teoria delle aspettative della struttura della scadenza" pp. 15

241. Christophe Croux, Mario Forni e Lucrezia Reichlin [1998] "A Measure of Comovement for Economic Indicators: Theory and Empirics"pp 23.

242. Carlo Alberto Magni [1998] "Note sparse sul dilemma del prigioniero (e non solo) pp 13.

243. Gian Paolo Caselli [1998] The future of mass consumption society in the former planned economies: a macro approach pp 21.

244. Mario Forni, Marc Hallin, Marco Lippi e Lucrezia Reichlin [1998] "The generalized dynamic factor model: identification and estimation pp 35.

245. Carlo Alberto Magni [1998] "Pictures, language and research: the case of finance and financial mathematics" pp 35.

246. Luigi Brighi [1998] "Demand and generalized monotonicity" pp 21.

247. Mario Forni e Lucrezia Reichlin [1998] "Risk and potential insurance in Europe" pp 20.

248. Tommaso Minerva, Sandra Paterlini e Irene Poli [1998] "A Genetic Algorithm for predictive Neural Network Design (GANND). A Financial Application" pp 12.

249. Gian Paolo Caselli Maurizio Battini [1998] "The Changing Distribution of Earnings in Poland from 1989 to 1996 pp. 9.

250. Mario Forni, Sergio Paba [1998] "Industrial Districts, Social Environment and Local Growth" Evidence from Italy pp. 27.

251. Lara Magnani [1998] "Un'analisi del distretto industriale fondata sulla moderna teoria economica dell'organizzazione" pp. 46.

252. Mario Forni, Lucrezia Reichlin [1998] "Federal Policies and Local Economies: Europe and the US" pp. 24.

253. Luigi Brighi [1998] "A Case of Optimal Regulation whit Multidimensional Private Information" pp 20.

254. Barbara Pistoresi, Stefania Luppi [1998] "Gli investimenti diretti esteri nell'America Latina e nel Sud Est Asiatico: 1982-1995" pp 27.

255 Paola Mengoli, Margherita Russo [1998] "Technical and Vocational Education and Training in Italy: Structure and Changes at National and Regional Level" pp 25.

256 Tindara Addabbo [1998] "On-the-Job Search a Microeconometric Analysis on Italian Data" pp. 29.

257 Lorenzo Bertucelli [1999] "Il paternalismo industriale: una discussione storiografica" pp.21.

258 Mario Forni e Marco Lippi [1999] "The generalized dynamic factor model: representation theory" pp. 25.

259 Andrea Ginzburg e Annamaria Simonazzi [1999] "Foreign debt cycles and the 'Gibson Paradox': an interpretative hypothesis" pp. 38.

260 Paolo Bosi [1999] "La riforma della spesa per assistenza dalla Commissione Onofri ad oggi: una valutazione in corso d'opera" pp. 56.

261 Marcello D'Amato e Barbara Pistoresi [1999] "Go and soothe the row. Delegation of monetary policy under private information" pp. 23.

262 Michele Lalla [1999] "Sampling, Maintenance, and Weighting Schemes for Longitudinal Surveys: a Case Study of the Textile and Clothing Industry" pp. 27.

263 Pederzoli Chiara e Torricelli Costanza [1999] "Una rassegna sui metodi di stima del Value at Risk (Var)".

264 Paolo Bosi, Maria Cecilia Guerra e Paolo Silvestri [1999] "La spesa sociale di Modena . La valutazione della condizione economica" pp 74.

265 Graziella Bertocchi e Michael Spagat [1999] "The Politics Co-optation" pp 14.

266 Giovanni Bonifati [1999] "The Capacity to Generate Investment. An analysis of the long-term determinants of investiment" pp.22.

267 Tindara Addabbo e Antonella Caiumi [1999] "Extended Income and Inequality by Gender in Italy" pp. 40.

268 Antonella Caiumi e Federico Perali [1999] "Children and Intrahousehold Distribution of Resources: An Estimate of the Sharing Rule of Italian Households" pp.24

269 Vincenzo Atella, Antonella Caiumi e Federico Perali [1999] "Una scala di equivalenza non vale l'altra" pp.23.

270 Tito Pietra e Paolo Siconolfi [1999] "Volume of Trade and Revelation of Information" pp. 33.

271 Antonella Picchio [1999] "La questione del lavoro non pagato nella produzione di servizi nel nucleo domestico (Household)" pp.58.

272 Margherita Russo [1999] "Complementary Innovations and Generative Relationships in a Small Business Production System: the Case of Kervit" pp.27.

273 André Dumas [1999] "L'Economie de la drouge" pp. 12.

274 Andrè Dumas [1999] "L'Euro à l'heure actuelle" pp. 12.

275 Michele Lalla Gisella Facchinetti [1999] "La valutazione dell'attività didattica: un confronto tra scale di misura e insiemi sfocati" pp.32.

276 Mario Biagioli [1999] "Formazione e valorizzazione del capitale umano: un'indagine sui paesi dell'Unione Europea" pp.21.

277 Mario Biagioli [1999] "Disoccupazione, formazione del capitale umano e determinazione dei salari individuali: un'indagine su microdati nei paesi dell'Unione Europea" pp.15.

278 Gian Paolo Caselli Giulia Bruni [1999] Il settore petrolifero russo, il petrolio del Mar Caspio e gli interessi geopolitici nell'area" pp. 28.

279 Luca Gambetti [1999] "The Real Effect of Monetary Policy: a New Var Identification Procedure" pp.22.

280 Marcello D'Amato Barbara Pistoresi [1999] "Assessing Potential Targets for Labour Market Reforms in Italy" pp. 8.

281 Gian Paolo Caselli Giulia Bruni e Francesco Pattarin [1999] "Gaddy and Ickes Model of Russian Barter Economy: Some Criticisms and Considerations" pp.10.

282 Silvia Muzzioli Costanza Torricelli [1999] "A Model for Pricing an Option with a Fuzzy Payoff " pp. 13.

283 Antonella Caiumi Federico Perali [1999] "Povertà e Welfare in Italia in Relazione alla Scelta della Scala di Equivalenza" pp.25.