Un algoritmo IML per la stima robusta dei modelli ARIMA e per l'individuazione dei valori anomali nelle serie storiche

Matteo Pelagatti
Facoltà di Scienze Statistiche
Università degli Studi di Milano Bicocca

Abstract

Chiunque si occupi di serie storiche sa bene quanto sia comune imbattersi in serie contaminate da valori anomali (spesso chiamati anche *outliers*). In una simile situazione è essenziale sia avere a disposizione strumenti per identificare e quantificare le anomalie, che contengono spesso informazioni importanti, sia predisporre modelli previsivi che "funzionino bene" anche in presenza di outliers. Del primo problema si occupano i metodi d'identificazione degli outliers e l'analisi d'intervento, mentre il secondo è oggetto della statistica robusta. La procedura che viene qui presentata ed implementata con SAS/IML assolve ai due compiti in modo coerente ed integrato.
Tra i modelli più utilizzati per l'analisi delle serie storiche ci sono sicuramente quelli della classe ARIMA, i cui parametri vengono stimati per mezzo del metodo della massima verosimiglianza (ML) o dei minimi quadrati (LS). Gli stimatori ML e LS, tuttavia, sono molto sensibili alla presenza di outliers, ed anche un'esigua frazione di valori anomali è sufficiente per inficiare le buone proprietà statistiche che li contraddistinguono. Ciononostante, e malgrado la presenza in letteratura di stimatori robusti per modelli ARMA (stimatori GM di ordine r, TRA, AM), in tutti i packages statistici più noti non esiste un'alternativa robusta agli stimatori ML e LS. La difficoltà d'implementazione di tali stimatori e la scarsa diffusione dei metodi robusti per serie storiche sono probabilmente le ragioni di tale deficienza. La procedura che verrà illustrata nei prossimi paragrafi fornisce stime robuste di modelli ARIMA e filtra la serie originale dai probabili valori anomali, in modo che la nuova serie, unitamente al modello ARIMA stimato, possa essere usata per avere previsioni attendibili. Tale procedura ha il vantaggio di essere facilmente implementabile, soprattutto in ambienti di sviluppo, come il SAS/IML, dotati di strumenti d'analisi numerica e calcolo matriciale. La facilità d'implementazione ha come contraltare la difficoltà di ricavare per via analitica le proprietà statistiche delle stime robuste fornite dalla procedura. Le conclusioni tratte sulla performance dello stimatore implementato nella procedura sono basate su una considerevole quantità di simulazioni e sono molto incoraggianti.

La nozione di robustezza

Il fine della stima robusta è quello di predisporre stimatori che (a) siano piuttosto efficienti sotto l'ipotesi di un modello centrale, e tali che (b) "moderati" cambiamenti nella distribuzione del campione su cui si basa la stima comportino solo "moderati" cambiamenti nella distribuzione dello stimatore. La formalizzazione del concetto espresso in (b), che prende il nome di *robustezza qualitativa*, si deve ad Hampel (1971)[1]. La definizione di Hampel, tuttavia, ha il limite di essere applicabile solamente ai casi in cui le osservazioni del campione sono i.i.d.[2]. Papantoni-Kazakos e Gray (1979), Bustos (1981), Cox (1981) e Boente, Fraiman e Yohai (1982) hanno fornito diverse generalizzazioni della definizione formale di robustezza qualitativa, utilizzabili nell'ambito dei processi stocastici.
Un concetto alternativo a quello di robustezza, ma ad esso collegato, è quello di *resistenza* (Tukey, 1976). Uno stimatore è resistente se è scarsamente influenzato (a) dalla presenza nel campione di "pochi" outliers e (b) dall'intervento di un "moderato" errore in tutte le osservazioni (per es. arrotondamenti). Mentre gli stimatori usuali (LS, ML) sono generalmente resistenti in quest'ultimo senso, raramente lo sono rispetto alla (a). Boente, Fraiman e Yohai, nel testo citato, formalizzano la relazione che intercorre tra robustezza qualitativa e resistenza.
Dato che uno stimatore robusto, sotto il modello teorico esatto, in genere non può essere più efficiente di uno stimatore di massima verosimiglianza, è opportuno affiancargli sempre una stima ML (o LS per processi gaussiani): se le due non differiscono molto, allora è probabile che il modello teorico sia correttamente specificato ed è quindi preferibile lo stimatore ML (o LS). In caso contrario lo stimatore robusto fornirà stime più affidabili.

Gli outliers nei modelli ARIMA

Si prendano in considerazione i processi stocastici rappresentabili attraverso la modellistica ARIMA

[1] Esistono varie definizioni di robustezza (minimax-robustness, B-robustness, V-robustness, robustezza d'efficienza, ecc.); la robustezza qualitativa è solitamente considerata il concetto di robustezza più forte ed importante.
[2] Indipendenti, identicamente distribuite.

$$\varphi(B)(z_t - \mu) = \theta(B)a_t, \quad (1)$$

dove $\{a_t\}$ è un *white noise* gaussiano con varianza σ_a^2,

$$\varphi(B) = 1 - \varphi_1 B - \ldots - \varphi_{p+d} B^{p+d} = \phi(B)\nabla^d = \\ = \left(1 - \phi_1 B - \ldots - \phi_p B^p\right)(1-B)^d \quad (2)$$

$$\theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q \quad (3)$$

ed i polinomi $\phi(B)$ e $\theta(B)$ hanno radici esterne al cerchio dell'unità.
Un processo ARIMA contaminato da outliers può essere formalizzato da

$$\{y_t\} = \{z_t\} + \{v_t\} \quad (4)$$

dove $\{z_t\}$ è un processo ARIMA e $\{v_t\}$ è un processo contaminante che con probabilità $(1-\varepsilon) > 0{,}5$ è uguale a zero[3]. Il valore ε è responsabile della quantità di outliers contenuti nella serie osservata e, leggendo il capitolo introduttivo del libro di Hampel et alt. (1986, sez. 1.2c), ci si può attendere che per dati di media qualità sia compreso tra 0,01 e 0,10. Il processo $\{v_t\}$ può essere specificato in modo tale da generare sia outliers isolati sia outliers raggruppati dotati di struttura dinamica. I casi considerati più di frequente nella letteratura sono due: *outliers innovativi* (IO) e *outliers additivi* (AO).

Outliers innovativi (IO).
Gli outliers innovativi non contaminano l'intera struttura dei modelli ARIMA, ma agiscono solamente sulla serie $\{a_t\}$ delle innovazioni (da cui il nome). Se nella (4) si pone

$$v_t = \mu + \frac{\theta(B)}{\varphi(B)}u_t$$

si ottiene

$$y_t = \mu + \frac{\theta(B)}{\varphi(B)}(a_t + u_t),$$

dove, per come si è definita v_t, $Pr(u_t = 0) = 1 - \varepsilon$.
In questa situazione, con l'ulteriore ipotesi $E\{(a_t + u_t)^2\} < \infty$, gli stimatori LS dei parametri φ_i e θ_j sono piuttosto stabili: infatti la loro matrice di covarianza asintotica è indipendente dalla distribuzione di $(a_t + u_t)$ (Whittle, 1962). Al contrario, la varianza dello stimatore LS di μ è proporzionale a $Var(a_t + u_t)$ ed è quindi piuttosto sensibile all'aumento della variabilità delle innovazioni. In presenza di IO è pertanto consigliabile applicare i minimi quadrati solo ai parametri ARMA, il livello della serie μ può essere stimato per mezzo di uno stimatore robusto di posizione (per es. uno stimatore M di posizione[4], o per semplicità la mediana). Nonostante la loro relativa stabilità, gli stimatori LS possono essere molto inefficienti rispetto ad altri tipi di stima. Una classe di stimatori robusti ed efficienti per modelli ARMA affetti da IO è quella degli *stimatori M* (*maximum-likelihood-type*) (si veda per esempio Martin e Yohai, 1985), per i quali predisporre un algoritmo di calcolo non presenta grosse difficoltà.
Gli IO, come si evince da quanto detto, non sono quasi mai un grave problema, purtroppo, tuttavia, sono outliers assai poco comuni nelle serie reali, specialmente se di tipo economico. Molto più frequenti e problematici da trattare sono gli outliers additivi.

Outliers additivi (AO).
Se nella (4) supponiamo che v_t abbia legge di distribuzione $\nu = (1-\varepsilon)\delta_0 + \varepsilon\mu$, dove δ_0 è una misura di probabilità degenere che pone tutta la sua massa in 0 ($Pr(\delta_0 = 0) = 1$), e μ è una legge di distribuzione solitamente ignota, allora il modello (4) genera un processo ARIMA contaminato da outliers additivi. Si può dimostrare (per es. Martin e Yohai, 1985) che in presenza di AO gli stimatori LS ed M sono distorti ed inefficienti. Per affrontare tale situazione sono state proposte diverse famiglie di stimatori: le classi dei *generalized M* (*GM*), degli *approximated M* (*AM*) e dei *residual autocovariance* (*RA*) sono le più note in letteratura[5]. Nel caso puramente autoregressivo gli stimatori GM ed RA hanno ottime proprietà di robustezza e sotto l'ipotesi di simmetria di *H* sono entrambi consistenti. I problemi sorgono quando si è in presenza della componente MA: in tal caso né i GM né gli RA forniscono stime qualitativamente robuste, pur rimanendo molto più stabili degli stimatori LS ed M. Masarotto (1987) e Bustos e Yohai (1986) hanno proposto versioni qua-

[3] La condizione $0 < \varepsilon < 0{,}5$ è necessaria per poter distinguere il processo contaminante da quello contaminato; infatti se più della metà delle osservazioni di una serie fossero outliers, sarebbe impossibile discernere le osservazioni "buone" da quelle "cattive".

[4] La stima M di posizione è data dal valore di μ per il quale $\Sigma\psi\left(\frac{x_i - \mu}{\sigma_x}\right) = 0$, con $\psi(z)$ funzione continua e tale che $sup|\psi(z)| < \infty$. Quando anche la varianza è incognita, nella formula si può usare $\hat{\sigma}_x = 1.483 \cdot MAD(x_i)$, dove MAD è la sigla di *median absolute deviation* (la mediana del valore assoluto degli scarti dalla mediana).

[5] Nell'articolo di Martin e Yohai (1985) vi è un'ottima rassegna degli stimatori citati.

litativamente robuste rispettivamente dei GM (GM di ordine *r*) e degli RA (RA tronchi o TRA); purtroppo l'implementazione di tali stimatori non risulta essere delle più semplici. Gli stimatori AM sono invece qualitativamente robusti per qualunque tipo di modello ARMA, tuttavia hanno una genesi piuttosto complessa ed una procedura di calcolo laboriosa.

La procedura che si sta per illustrare nasce dalla convinzione che uno stimatore, affinché sia utilizzabile, oltre ad avere buone proprietà statistiche, debba essere calcolabile senza troppi problemi: tanto più che uno studioso che utilizza strumenti statistici non necessariamente è uno statistico. Dato che gli stimatori robusti di cui si è appena scritto non sono implementati in alcun pacchetto statistico tra quelli più noti, si è voluta studiare una procedura che, oltre a fornire stime robuste con buone proprietà statistiche (verificate solo empiricamente), fosse traducibile in un algoritmo di calcolo senza grosse difficoltà.

La procedura: introduzione

Hampel in uno scritto del 1974 introdusse un fondamentale strumento della statistica robusta: la *curva d'influenza*, poi ribattezzata *funzione d'influenza* (*IF*). Per una completa trattazione dell'approccio alla statistica robusta basato sulla funzione d'influenza si rimanda al già citato libro di Hampel et alt. (1986), in questa sede è sufficiente dire che tale funzione (nella sua versione finita) descrive il "potere" di modificare una stima che una singola osservazione aggiuntiva ha a seconda del suo valore. Per mezzo della IF si è potuto constatare che le stime LS sono in genere molto sensibili alle osservazioni i cui valori giacciono sulle code della distribuzione da cui provengono (per es. l'influenza di una osservazione aggiuntiva sulla media campionaria è proporzionale alla distanza del valore dell'osservazione dalla media campionaria precedentemente calcolata). La generalizzazione della funzione d'influenza all'ambito delle le serie storiche è stata prodotta da Martin e Yohai (1984), i quali hanno anche calcolato la IF dello stimatore LS dei parametri dei modelli ARMA: i loro risultati confermano l'alta sensibilità delle stime LS alle osservazioni estreme.
Compiendo qualche semplificazione, si può dire che un modo per ricavare stimatori robusti è quello di modificare uno stimatore esistente in modo da limitare l'influenza, che un'osservazione estrema può avere sulla stima. Gli stimatori GM di cui si è sopra accennato nascono, nell'ambito della regressione, proprio nell'ottica appena descritta (da cui il nome alternativo *bounded-influence estimates*): se si suppone che gli errori di regressione del modello centrale siano normali, lo stimatore GM è uno stimatore ai minimi quadrati ponderati,

$$\sum w_i \left(y_i - \mathbf{b}'\mathbf{x}_i\right)\mathbf{x}_i = 0 \qquad (5)$$

dove $\mathbf{x}_i$ è il vettore delle variabili esplicative, $\mathbf{b}$ il vettore dei coefficienti di regressione incogniti ed i pesi w_i sono calcolati in modo che nessun addendo possa dominare la sommatoria (si rimanda al libro di Hampel et alt. citato per la spiegazione del calcolo dei pesi). Quando i GM vengono applicati ai modelli AR, essi mantengono le proprietà di robustezza che possiedono nel caso regressivo; quando invece nel modello è presente la componente MA, un singolo valore anomalo al tempo t_0 si ripercuote sulla stima dei residui ai tempi $t_0, t_0+1, t_0+2, \ldots, n$. Dato che le stime dei parametri ARMA sono ottenute come funzione dei white noise stimati, i loro valori saranno pertanto a loro volta influenzati. Per mostrare quanto appena detto si prenda come esempio il processo MA(1), $y_t = a_t - \theta a_{t-1}$. I residui stimati $\hat{a}_t(\theta)$ possono essere calcolati con la formula

$$\hat{a}_t(\theta) = y_t + \theta y_{t-1} + \theta^2 y_{t-2} \ldots + \theta^{t-1} y_1 , \qquad (6)$$

per cui basta che un valore precedente o contemporaneo a *t* sia anomalo perché tutte le stime dei valori del white noise da *t* in poi siano in qualche modo compromesse: l'influenza dell'outlier su $\hat{a}_t(\theta)$ è tanto più grande quanto il processo si avvicina alla zona di non invertibilità (nell'esempio quando θ è prossimo ad uno).
Anche alla base della procedura che si sta per illustrare vi sono i minimi quadrati ed il contenimento dell'influenza che le osservazioni estreme hanno sulle stime dei parametri; in più si è posto rimedio al problema di stima del white noise che si presenta quando il processo comprende una componente MA (e la serie contiene outliers additivi). Per illustrare intuitivamente come funziona la procedura, si supponga di avere una serie $\{y_t\}$ generata da un processo ARMA(1,1) gaussiano a media nulla, $y_t = \phi y_{t-1} + a_t - \theta a_{t-1}$, contaminato da qualche AO in posizioni sconosciute. Si supponga anche di conoscere il valore della deviazione standard del white noise o di una sua stima. La stima LS dei parametri del modello in esame è data da quei valori di ϕ e θ per cui vale

$$\sum_{t=2}^{n} \left(y_t - \phi y_{t-1} + \theta \hat{a}_{t-1}\right)^2 = \min$$

con (7)

$$\begin{cases} \hat{a}_t = y_t - \phi y_{t-1} + \theta \hat{a}_{t-1} & t \geq 2 \\ \hat{a}_t = 0 & t < 2 \end{cases}$$

Per evitare che un addendo possa dominare la sommatoria nella (7) è possibile porre un limite al valore

massimo che $|\hat{a}_t|$ e $|y_t|$ possono assumere, cercando, tuttavia, di non "alterare troppo" il comportamento della stima quando la serie è priva di outliers. Gli stimatori GM raggiungono l'obiettivo diminuendo il peso degli addendi che, secondo alcuni criteri (per es. Martin e Yohai, 1985), risultano "troppo grandi", ma hanno il limite, appena esposto, di non essere robusti in presenza di componenti MA. Osservando la (6) sembra naturale ottenere la robustezza sostituendo le osservazioni sospettate di essere outliers con dei valori che causino un minor danno alla stima: se vi fosse la certezza che una certa osservazione è anomala sarebbe opportuno sostituirla con la sua migliore previsione (nel caso dell'esempio data da $\hat{y}_t = \phi y_{t-1} - \theta \hat{a}_{t-1}$); quando invece vi è dubbio sulla "bontà" di una osservazione è più utile modificarla in modo tale che il cambiamento, se essa è genuina, non induca un grosso effetto sulla stima, ma al contempo preservi quest'ultima dagli effetti nefasti degli outliers. Riprendendo il modello ARMA(1,1) esemplificativo si illustrerà come chi scrive ha pensato di mettere in pratica quanto appena descritto. Si supponga per il momento di conoscere i valori di ϕ e θ; dato che l'errore di previsione

$$a_t = y_t - \hat{y}_t = y_t - \phi y_{t-1} + \theta a_{t-1}$$

è, per ipotesi, un processo white noise gaussiano, a_t/σ_a ha distribuzione normale standardizzata. Un valore di a_t/σ_a sulle code della distribuzione N(0,1) candiderà l'osservazione contemporanea y_t ad essere un possibile outlier. Per esempio si fissi l'evento $|a_t/\sigma_a| > 2{,}576$ (cui corrisponde una probabilità pari a 0,01), e si supponga che calcolando su una serie reale $\hat{a}_t = y_t - \hat{y}_t$, l'evento fissato si verifichi per la 10 osservazione (t=10). Sostituendo il valore calcolato di $\hat{a}_{10}$ con

$$2.576 \cdot \sigma_a \cdot \text{sgn}(\hat{a}_{10})$$

si pone un limite al valore massimo di $|\hat{a}_{10}|$, limitandone l'influenza nella (7). Non avendo, a questo punto, la certezza della natura di y_{10} $(= \hat{y}_{10} + \hat{a}_{10})$, attribuendogli il valore

$$y_{10} = \hat{y}_{10} + 2.576 \cdot \sigma_a \cdot \text{sgn}(\hat{a}_t),$$

coerente col nuovo residuo, si pone un limite all'influenza che y_{10} ha sia nella (7), sia nel calcolo dei residui di previsione successivi. Se si compiono sostituzioni analoghe a quelle appena descritte per tutte le osservazioni per le quali vale $|a_t/\sigma_a| > 2{,}576$, si ottiene una serie *filtrata*, uguale a quella originaria tranne in quei casi in cui si è operata la sostituzione (nell'esempio, dato che quando la serie non è contaminata, $\Pr(|a_t/\sigma_a| > 2{,}576) = 0{,}01$, le osservazioni modificate sono approssimativamente l'1%[6]).

Dato che il fine della procedura è la stima dei parametri del modello ARIMA, si lasci cadere l'ipotesi fatta per la quale ϕ e θ (dell'esempio) erano supposti noti. È ovvio che ignorando i valori dei parametri non è possibile calcolare le previsioni $\hat{y}_t$, necessarie per apportare le correzioni alla serie appena descritte. In situazioni analoghe solitamente si ricorre ad una stima LS preliminare dei parametri per applicare in modo iterativo la "correzione" della serie ed i minimi quadrati della serie modificata; tuttavia la convergenza non sempre è garantita. Nella sezione che segue è contenuta, oltre che la descrizione formale della procedura di stima, una soluzione al problema alternativa a quella iterativa ora accennata.

La procedura: stima dei parametri ARIMA

Sia

$$\varphi(B)(x_t - \mu) = \theta(B)a_t \tag{8}$$

il modello ARIMA(p,d,q) che si desidera stimare.

Si supponga di avere una stima $\hat{\sigma}_a^2$ della varianza del white noise $\{a_t\}$ (un modo per ottenere tale stima sarà descritto nella prossima sezione). Per brevità di notazione si ponga $y_t = x_t - \mu$.

Le stime robuste dei parametri del modello, che qui si propongono, sono date da quei valori di $(\mu, \varphi_1, ..., \varphi_{p+d}, \theta_1, ..., \theta_q)$ per cui vale

$$\sum_{t=p+d+1}^{n} \tilde{a}_t^2 = \min, \tag{9}$$

dove:

$$\begin{cases} \tilde{a}_t = 0 & t \le p+d \\ \tilde{a}_t = \hat{\sigma}_t \cdot f\left(\dfrac{y_t - \hat{y}_t}{\hat{\sigma}_t}\right) & t > p+d \end{cases}, \tag{10}$$

è una serie di residui di previsione modificata per mezzo della funzione f, di cui si dirà;

[6] In realtà, con $k = 2{,}576$, il numero atteso di osservazioni modificate non sarà proprio l'1%, perché la modifica anche di un solo valore della serie si ripercuote sulla stima di tutti i residui di previsione; tuttavia per valori di $z_{\alpha/2}$ (di una normale standardizzata) per cui α è sufficientemente piccolo, l'approssimazione risulta soddisfacente (ai nostri fini).

$$\hat{y}_t = \varphi_1 \tilde{y}_{t-1} + ... + \varphi_{p+d} \tilde{y}_{t-p-d} + \\ -\theta_1 \tilde{a}_{t-1} - ... - \theta_q \tilde{a}_{t-q} \quad , \qquad (11)$$

$$\text{con} \quad t > p + d$$

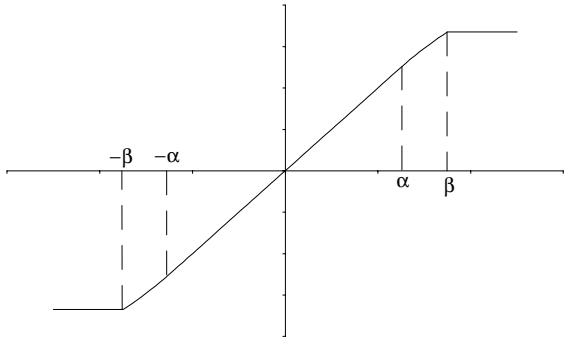
è una serie di previsioni basata sui residui $\{\tilde{a}_t\}$ e sulla serie filtrata

$$\begin{cases} \tilde{y}_t = y_t & t \le p + d \\ \tilde{y}_t = \hat{y}_t + \tilde{a}_t & t > p + d \end{cases} . \qquad (12)$$

La funzione *f* che verrà utilizzata è:

$$f(x) = \begin{cases} x & |x| \le \alpha \\ \text{sgn}(x)\sqrt{2\alpha|x| - \alpha^2} & \alpha < |x| \le \beta \\ \text{sgn}(x)\sqrt{2\alpha\beta - \alpha^2} & \beta < |x| \end{cases} , \qquad (13)$$

il cui grafico è:



Le costanti α e β, che possono anche coincidere, sono responsabili della robustezza della stima: quando $\alpha = \beta = +\infty$, lo stimatore proposto coincide con lo stimatore LS, infatti, come è facile ricavare dalle formule (10)-(13), $\{\tilde{a}_t\}$ diventa l'usuale serie dei residui, $\{\hat{y}_t\}$ coincide con la serie delle proiezioni lineari *one-step-ahead* e $\{\tilde{y}_t\} = \{y_t\}$. Per processi ARIMA gaussiani, funzionano bene valori di α e β nell'intervallo [1.960, 3.291], che corrispondono ad una densità delle code della normale pari rispettivamente a 0.05 e 0.001. Più α e β sono grandi, più lo stimatore è efficiente sotto il modello esatto ipotizzato e meno è robusto rispetto alla presenza di outliers.
Formalizzato nel modo appena visto, lo stimatore potrebbe sembrare, al contrario di quanto anticipato, piuttosto poco intuitivo; tuttavia, se descritto sotto forma di procedura, lo stesso stimatore acquisisce notevole chiarezza.

Procedura I.

Per semplicità si illustrerà la procedura per $\alpha = \beta$.
Per mezzo di un metodo di ottimizzazione numerico (IML ne implementa diversi), si trovi il valore del vettore di parametri $(\mu, \varphi_1, ..., \varphi_{p+d}, \theta_1, ..., \theta_q)$ che minimizza la funzione LOSS così definita:

LOSS $(\mu, \varphi_1, ..., \varphi_{p+d}, \theta_1, ..., \theta_q)$

Sia $t = p + d + 1$.

1. si calcoli l'usuale proiezione *one-step-ahead* ($\hat{y}_t$)
2. si calcoli l'usuale residuo di proiezione $\tilde{a}_t = y_t - \hat{y}_t$;
3. se $|\tilde{a}_t| \le \alpha\hat{\sigma}_a$ e $t < n$ si torni al passo 1. incrementando il valore di *t*, se $t = n$ si vada al passo 6.
4. si modifichi il valore del residuo: $\tilde{a}_t = \text{sgn}(\tilde{a}_t)\alpha\hat{\sigma}_a$;
5. si modifichi il valore della serie in coerenza col nuovo residuo: $y_t = \hat{y}_t + \tilde{a}_t$; se $t < n$ si torni al passo 1;
6. si restituisca la somma $\sum_{t=p+d+1}^{n} \tilde{a}_t^2$ come valore della funzione LOSS.

Per ricondurre la procedura I. alla formalizzazione (9)-(13) si noti che sono stati utilizzati gli stessi simboli per tutto tranne che per la serie $\{\tilde{y}_t\}$, che non compare esplicitamente nella procedura, ma che corrisponde alla serie $\{y_t\}$ dopo le modifiche del passo 5. La funzione (13), nel caso considerato in cui $\alpha = \beta$, è implementata nei passi 3. e 4.
Per motivi di ordine computazionale è conveniente stimare il livello della serie μ per mezzo di uno stimatore robusto di posizione, quale uno stimatore M o, per semplicità, la mediana, piuttosto che includere μ tra gli argomenti della funzione LOSS.

La procedura: stima congiunta dei parametri ARIMA e della varianza del white noise

Per rendere utilizzabile la procedura descritta nella sezione precedente è una stima della deviazione standard del white noise che compare nella (10). L'algoritmo che viene ora presentato integra la procedura della sezione precedente con una stima congiunta della deviazione standard del white noise.

Procedura II.

1. Stima preliminare LS del modello e stima preliminare di σ_a per mezzo di della *median absolute deviation* (MAD) dei residui moltiplicata per il fattore di correzione 1.483 (cfr. nota 4).
2. Calcolo dei parametri ARIMA per mezzo della procedura I. e dell'ultima stima di σ_a
3. Nuova stima di σ_a come 1.483·MAD dei residui $\{\tilde{a}_t\}$ ottenuti nel punto 2.

4. Se $\left|\hat{\sigma}_a^{(i)} - \hat{\sigma}_a^{(i-1)}\right| \geq \varepsilon$, dove l'apice ($i$) indica a quale iterazione si riferisce la stima $\hat{\sigma}_a$ ed ε è una soglia di tolleranza prestabilita, allora si ritorna al passo 2. per una nuova iterazione, altrimenti si termina la procedura e si utilizzano le stime dei parametri e della deviazione standard ottenute nell'ultima iterazione (i).

Per mezzo della Procedura II. si ottengono quindi stime robuste di tutti i parametri di un modello ARIMA. Tuttavia, oltre alla stima dei parametri, la procedura fornisce una serie ($\{\tilde{y}_t\}$ della (12)) corretta rispetto alla presenza di valori che sono estremi nel modello ARIMA stimato. Quest'ultima serie può esser utilizzata in luogo della serie originale $\{y_t\}$ per fini previsivi; infatti, anche ammesso di avere a disposizione i "veri" parametri del modello ARIMA, le previsioni ottenute da una serie contaminata risultano comunque compromesse, specialmente quando gli outliers si trovano nei dati più recenti. La serie $\{\tilde{y}_t\}$ è preferibile alla serie originale anche quando interessi il periodogramma o la densità spettrale della serie, e questa sia affetta da outliers: infatti l'influenza di valori anomali sulla trasformata discreta di Fourier (DFT) è notevole (si ricordi che i valori della DFT possono essere visti come stime LS dei coefficienti di una combinazione lineare di seni e coseni). La serie $(y_t - \tilde{y}_t)_{t=1,\ldots,n}$, che è nulla per la maggior parte dei t, è uno strumento utile per individuare la posizione degli outliers nella serie, ed anche per valutare quanto *outlying* essi siano: i valori diversi da zero indicano la presenza di outliers e più il loro modulo è alto, più è probabile che il valore provenga da un processo contaminante. Quest'ultimo strumento è particolarmente utile quando si voglia applicare l'analisi d'intervento (Box e Tiao, 1975) alla serie in esame.

Simulazioni

Si riportano i risultati di alcune delle simulazioni eseguite. L'algoritmo utilizzato per effettuare le simulazioni è quello riportato in appendice. Con esso si sono calcolate sia le stime robuste (RB) sia quelle LS. Le sintesi contenute nelle tabelle che seguono si riferiscono a gruppi di 500 serie di 200 osservazioni generate con la funzione ARMASIM di IML. Si sono generate 500 serie secondo un modello ARMA(1,1) con $\phi = .5$, $\theta = -.8$ e $\sigma_a = 10$. Per generare le 500 serie contaminate si sono sommate alle serie prodotte nel modo descritto delle serie il cui generico termine fosse nullo con probabilità .95 e fosse con probabilità .05 realizzazione di una normale con media nulla e deviazione standard pari a tre volte quella delle serie.

Nelle tabelle si sono riportate: la media dei parametri stimati per le 500 serie (media), lo scarto della media rispetto al valore reale del parametro (dist.), l'errore quadratico medio delle stime (eqm), e l'efficienza relativa rispetto alle stime LS (eff. rel.).

Processo incontaminato. $\alpha = 2.576$ $\beta = 3$

	$\hat{\phi}_{LS}$	$\hat{\theta}_{LS}$	$\hat{\phi}_{RB}$	$\hat{\theta}_{RB}$
media	0.5079	-0.7954	0.5073	-0.7997
dist.	0.0079	0.0046	0.0073	0.0003
eqm	0.0041	0.0026	0.0043	0.0027
eff. rel.	1.0000	1.0000	0.9602	0.9472

Processo contaminato. $\alpha = 2.576$ $\beta = 3$

	$\hat{\phi}_{LS}$	$\hat{\theta}_{LS}$	$\hat{\phi}_{RB}$	$\hat{\theta}_{RB}$
media	0.5014	-0.0529	0.5354	-0.4340
dist.	0.0014	0.7471	0.0354	0.3660
eqm	0.0138	0.5912	0.0102	0.1602
eff. rel.	1.0000	1.0000	1.3565	3.6899

$\alpha = 2.576$ $\beta = 3$

	Pr. incontaminato		Pr. contaminato	
	$\hat{\sigma}_{LS}$	$\hat{\sigma}_{RB}$	$\hat{\sigma}_{LS}$	$\hat{\sigma}_{RB}$
media	9.98	9.96	18.16	12.02
dist.	-0.02	-0.04	8.16	2.02
eqm	0.22	0.63	73.48	5.92
eff. rel.	1.00	0.35	1.00	12.41

Processo incontaminato. $\alpha = 2.576$ $\beta = 2.576$

	$\hat{\phi}_{LS}$	$\hat{\theta}_{LS}$	$\hat{\phi}_{RB}$	$\hat{\theta}_{RB}$
media	0.4924	-0.7988	0.4912	-0.8080
dist.	-0.0076	0.0012	-0.0088	-0.0080
eqm	0.0046	0.0025	0.0049	0.0031
eff. rel.	1.0000	1.0000	0.9485	0.8000

Processo contaminato. $\alpha = 2.576$ $\beta = 2.576$

	$\hat{\phi}_{LS}$	$\hat{\theta}_{LS}$	$\hat{\phi}_{RB}$	$\hat{\theta}_{RB}$
media	0.4785	-0.0803	0.5167	-0.5338
dist.	-0.0215	0.7197	0.0167	0.2662
eqm	0.0155	0.5553	0.0090	0.0933
eff. rel.	1.0000	1.0000	1.7278	5.9544

$\alpha = 2.576$ $\beta = 3$

	Pr. incontaminato		Pr. contaminato	
	$\hat{\sigma}_{LS}$	$\hat{\sigma}_{RB}$	$\hat{\sigma}_{LS}$	$\hat{\sigma}_{RB}$
media	9.95	9.93	17.92	11.62
dist.	-0.05	-0.07	7.92	1.62
eqm	0.27	0.67	70.51	4.21
eff. rel.	1.00	0.40	1.00	16.75

Appendice: funzioni IML

Sono di seguito riportate le funzioni in codice IML utilizzate per il calcolo dello stimatore proposto. La funzione principale, è ROBARMA, che chiama MAD, LOSS e quest'ultima chiama a sua volta SQRT_RHO. MAD calcola la median absolute deviation, SQRT_RHO calcola la funzione (13), LOSS implementa la procedura I. e ROBARMA implementa la procedura II.
Le funzioni MAD e SQRT_RHO utilizzano solo variabili locali.

```
/*
FUNCTION NAME   : MAD
ARGUMENT        : x column vector of data
RETURNS: scalar value of MAD
LOCAL VARIABLES  : y median-centered series
*/
START MAD(x);
  y = x - MEDIAN(x);
  RETURN(MEDIAN(ABS(y)));
FINISH MAD;

/*
FUNCTION NAME   : SQRT_RHO
ARGUMENT        : x       scalar
                  : alpha   tuning constant
                  : beta    tuning constant
                  : sigma   standard deviation of x
RETURNS: value of robustifying function
*/
START SQRT_RHO(x,alpha,beta,sigma);
  y=x/sigma;
  IF ABS(y)<=alpha THEN z=y;
  ELSE DO;
     IF ABS(y)<=beta THEN
     z=SQRT(2#alpha#ABS(y)-alpha##2)#y/ABS(y);
          ELSE
          z=SQRT(2#alpha#beta-alpha##2)#y/ABS(y);
  END;
  RETURN(z#sigma);
FINISH SQRT_RHO;
```

Le funzioni ROBARMA e LOSS condividono le seguenti variabili globali (tutte le variabili globali hanno nome preceduto dal carattere _):

_wseries	serie di lavoro, centrata con la mediana
_n	numero di osservazioni nella serie
_p	ordine AR massimo
_q	ordine MA massimo
_zerone	vettore $(_p + _q + 1)$x1 di 0 e 1 indicante quali parametri AR, MA e di posizione sono da stimare
_sigma	stima robusta della deviazione standard del white noise
_alpha	costante α della (13)
_beta	costante β della (13)
_ls	vettore 1x$(_p + _q)$ contenete le stime LS
_sigmals	stima LS della deviazione standard
_filt	serie corretta $\{\tilde{y}_t\}$
_resid	serie dei residui di previsione

```
/*
FUNCTION NAME   : LOSS
ARGUMENT        : param 1x(_p+_q) vector of parame-
                  ters (then redifined as col. vector)
RETURNS         : scalar value of LOSS function
LOCAL VARIABLES : fitted (100+_n)x1 vector one-step-
                  ahead forecasts
                  : resid (100+_n)x1 vector residuals
                  w.r. to fitted
                  : lseries   (100+_n)x1 vector local work
                  series (must be 0-centered)
*/
START LOSS(param)
GLOBAL(_wseries,_n,_p,_q,_zerone,_sigma,_alpha,_beta,
_filt,_resid);
  * LOCAL VARIABLES ASSIGNMENT;
  param = param`;
  fitted=J(100+_n,1,0);
  resid=J(100+_n,1,0);
  lseries={[100] 0}` // _wseries;
  * LOOPS TO COMPUTE ONE-STEP-AHEAD FORECASTS;
  DO i = (101+_p) TO (100+_n) BY 1;
          /* AR LOOP */
          DO j = 1 TO _p BY 1;
                  fitted[i,1] = fitted[i,1] + param[j,1] #
                  lseries[i-j,1];
          END;
          /* MA LOOP */
          DO h = 1 TO _q BY 1;
                  fitted[i,1] = fitted[i,1] - param[_p+h,1]
                  # resid[i-h,1];
          END;
          /* RESIDUAL AND OBSERVATION CORRECTIONS
          WITH A ROBUSTIFYING FUNCTION */
          resid[i,1] = SQRT_RHO(lseries[i,1] - fit-
          ted[i,1],_alpha,_beta,_sigma);
          lseries[i,1] = fitted[i,1] + resid[i,1];
  END;
  * ASSIGNES GLOBALS;
  _filt = lseries[101:100+_n,1];
  _resid = resid[101:100+_n,1];
  * RETURNS THE SUM OF THE SQUARED RESIDUALS;
  RETURN(resid` * resid);
FINISH LOSS;


/*
FUNCTION NAME   : ROBARMA
ARGUMENTS       : series column vector of series
                  : p scalar max order of AR
                  : q scalar max order of MA
                  : zerone (p+q+1)x1 vector of parame-
                  ter to estimate:
                  ar(1) ... ar(p) ma(1) ... ma(q) location
                  0 - 1 ... 0 - 1 0 - 1     0 - 1 0 - 1
                  1 to estimate, 0 otherwise,
                  : alpha 1st tuning constant for robust
                  function
                  : beta 2nd tuning constant for robust
                  function
LOCAL VARIABLES
                  : param   1x(p+q) vector of estimated
                  ARMA parameters
                  : wparam 1x(p+q) work vector for e-
                  stimated ARMA parameters
                  : zeros
                  : lc

*/
```

```
START ROBARMA(series,p,q,zerone,alpha,beta)
GLOBAL(_wseries,_n,_p,_q,_zerone,_sigma,_alpha,_beta,
_filt,_fit,_resid,_ls, _sigmals);
  * GLOBAL VARIABLES ASSIGNMENTS;
  _wseries = series - zerone[p+q+1,1] # MEDIAN(series);
  _n = NROW(series);
  _p = p;
  _q = q;
  _zerone = zerone;
  _alpha = 1000000000;
  /* _alpha AND _beta INITIALIZED IN ORDER TO FIRST
  LS-ESTIMATE ARMA PARAMETERS */
  _beta = 1000000000;
  _sigma = 1;
  /* _sigma IS INITIALIZED TO 1 AS 1st ITERATION IS
  LS,AND _sigma-ESTIMATE IS NOT NEEDED YET */
  * LOCAL VARIABLES ASSIGNMENTS;
  param = J(1,p+q,0.001);
  /* STARTING VALUE FOR ESTIMATION */
  * CREATION OF THE LINEAR CONSTRAINS MATRIX FOR
  THE NLP* FUNCTION CALL;
  zeros = LOC(^zerone[1:p+q,1]);
  /* zeros IS A ROW VECTOR CONTAINING INDEXES OF
  ZERO VALUED ELEMENTS IN zerone EXCEPT THE
  LOCATION PARAMETER */
  lc = J(2+NCOL(zeros),p+q+2,0);
  /* lc IS THE LINEAR CONSTRAINS MATRIX, INITIALIZED
  WITH 0's */
  lc[1:2,1:p+q+2]=.; /* FIRST 2 ROWS ARE SET TO .'S */
  DO i=1 TO NCOL(zeros) BY 1;
          lc[2+i,zeros[i]] = 1;
  END;
  * LS ESTIMATES IN ORDER TO HAVE A FIRST ESTIMATE
  OF _sigma WITH 1.483MAD;
  CALL NLPNMS(rc,_ls,"LOSS",param,,lc);
  _sigmals=SQRT((_resid`*_resid)/_n);
  _sigma = 1.483#MAD(_resid);
  param[1,1:p+q]=_ls[1,1:p+q];
  /* ONLY LS-ESTIMATES OF AR PARS. ARE USED AS
  STARTING POINT FOR ROBUST ESTIMATES AS LS-
  ESTIMATES OF MA PARS. CAN BE EXTREMELY BIASED */
  * ROBUST ESTIMATION;
  _alpha = alpha;
  _beta = beta;
  count=0;
  /* COUNTER INITIALIZATION */
  diff=1000;
  /* TOLLERANCE VARIABLE INITIALIZATION */
  DO UNTIL(count>=50|diff<0.01);
  /* MAX ITERATIONS=50 AND TOLLERANCE(ON
  _sigma)=0.01 */
  count=count+1;
  CALL NLPNMS(rc,xr,"LOSS",param,,lc);
  param=xr;
  diff=_sigma - 1.483#MAD(_resid);
  _sigma = 1.483#MAD(_resid);
  END;
  RETURN(param);
FINISH ROBARMA;
```

Bibliografia

Boente G., Freiman R. and Yohai V. (1982). “Qualitative robustness for general stochastic processes”, *Technical Report 26*, Univ. Of Washington, Dept. Of Statistics.

Bustos O. H. (1981). “Qualitative robustness for general stochastic processes”, *Informes de Matematica*, Serie B-002/81, Insituto de Matematica Pura e Aplicada, Brazil.

Bustos O. H. , Yohai V. J. (1986) - "Robust Estimates for ARMA Models", *JASA*, 81, 155-168.

Cox D. (1981). “Metric on stochastic processes and qualitative robustness”, *Technical Report 3*, Univ. Of Washington, Dept. Of Statistics.

Hampel F.R. (1971). “A general qualitative definition of robustness”, *Ann. Math. Statist.*, 42,1887-1896.

Hampel F.R. (1974). “The influence curve and its role in robust estimation”, *JASA*, 69, 383-393.

Hampel F.R., Rousseeuw P.J. and Ronchetti, E. (1981). “The change-of-variance curve and optimal redescending *M*-estimators”, JASA, 76, 643-648.

Martin R. D., Yohai V. J. (1985) - “Robustness in Time Series and Estimating ARMA models”, in Hannan E. J., Krishnaiah P. R., Rao M. M. (Eds), *Handbook of Statistics*, Vol. 5. Elsevier Science Publishers, pp. 119-155.

Masarotto G. (1987) - "Robust and Consistent Estimates of Autoregressive Moving Average Parameters", *Biometrika*, 74, 791-797.

Papantoni-Kazakos P. e Gray R. M. (1979). “Robustness of estimators on stationary observations”, Ann. Probab. 7, 989-1002.

Tukey J.W. (1976) - “Useable resistant/robust techniques of analysis”, in: Nicholson and Harris, eds., Proc. First ERDA Statistics Symposium. Batelle Northwest Laboratories, Richland, WA.

Whittle, P. (1962) - “Gaussian estimation in stationary time series”, *Bull. Int. Statist.* 39, 105-129.