

Calcolo delle regressioni non lineari. Metodi lineari generalizzati (*)

Agostino Tarsitano
Università degli studi della Calabria
Dipartimento di economia e Statistica
87030 Arcavacata di Rende (Cs)
agotar@unical.it

Riassunto.

In questo lavoro si esaminerà la soluzione dei sistemi di equazioni non lineari con delle procedure che sorgono come una naturale estensione dei metodi iterativi impiegati per i sistemi di equazioni lineari. In particolare si approfondirà la combinazione del metodo non lineare di Newton con i metodi lineari Jacobi e Gauss-Seidel. Il contesto analitico a cui si farà più immediato riferimento è la soluzione diretta del sistema delle equazioni normali per la determinazione degli stimatori ai minimi quadrati.

Keywords:

regressione non lineari, procedura Jacobi, procedura Gauss-Seidel.

() Dipartimento di Economia politica, 1980.*

1. Il contesto analitico

Consideriamo il problema di determinare, con il metodo dei minimi quadrati, il vettore dei parametri $\beta=(\beta_1, \beta_2, \dots, \beta_k)$ che sia il minimo assoluto, in un qualche sottoinsieme di R^k , per la somma dei residui quadrati:

$$S(\beta) = \sum_{i=1}^n [y_i - f(X_i; \beta)]^2 \quad (1.1)$$

dove le $\{y_i\}$ sono valori osservati sulla variabile endogena di un dato sistema e le $f(X_i, \beta)$ sono i valori realizzati nella corrispondente funzione di risposta in dipendenza dei parametri β e dei regressori X_i . Al fine di semplificare la trattazione analitica del problema facciamo le seguenti ipotesi:

- 1) $\beta \in B \subset R^k$ con B compatto e convesso.
- 2) $S(\beta)$ ha un unico punto di minimo β^* nell'insieme B . Inoltre, β^* è interno a B .
- 3) Per ogni valore dell'indice i le $f(X_i, \beta)$ sono dotate di derivate prime e seconde rispetto a β continue su B .
- 4) Se indichiamo con $H(\beta)=q'(\beta)$ l' Hessiano della (1.1) questo sarà in B sempre positivo definito.
- 5) Si dispone di una stima iniziale dei parametri β^0 , pure appartenente B .

Le ipotesi (1-5) afferiscono alla determinazione analitica e numerica del valore di β^* mentre, per le proprietà statistiche che assicurino che β^* sia il valore di uno stimatore efficiente e dotato delle usuali proprietà asintotiche si possono vedere Jenrich (1969) e Malinvaud (1970).

Le condizioni esposte incorporano, nella loro drasticità, una accurata ricerca preliminare sul comportamento di $S(\beta)$ nello spazio parametrico in modo da poter circoscrivere il sottospazio B all'interno del quale possano valere le condizioni richieste. La ricerca avrà soprattutto lo scopo di localizzare approssimativamente β^* determinare β^0 , e controllare che β^0 sia abbastanza isolato cioè non troppo vicino ad altri punti critici di $S(\beta)$ la eventualmente presenti nello spazio parametrico che potrebbero rendere fallaci molti metodi iterativi anche con valori iniziali apprezzabili (Tarsitano, 1979).

La determinazione del punto di minimo della funzione (1.1) può essere effettuata, nelle condizioni poste, risolvendo le equazioni corrispondenti alle condizioni del primo ordine. Indichiamo con $q(\beta)$ il vettore delle k derivate parziali prime della $S(\beta)$ rispetto a β^*

$$q_i(\beta) = q_i(\beta_1, \beta_2, \dots, \beta_k) = \frac{\partial S(\beta)}{\partial \beta_i}, \quad i = 1, 2, \dots, k \quad (1.2)$$

Le ipotesi (1-5) assicurano che il punto di minimo assoluto esiste ed è dato dall'unica soluzione β^* del sistema

$$\left\{ \begin{array}{l} q_1(\beta_1, \beta_2, \dots, \beta_k) = 0 \\ q_2(\beta_1, \beta_2, \dots, \beta_k) = 0 \\ \dots \quad \dots \\ q_k(\beta_1, \beta_2, \dots, \beta_k) = 0 \end{array} \right. \quad (1.3)$$

I metodi iterativi esaminati in questo articolo sono orientati alla soluzione del sistema (1.3). Tale approccio alle regressioni non lineari è molto meno praticato rispetto alle procedure (cfr. Goldfeld et al. 1966; Hartley, 1961; Marquardt, 1963) che perseguono la minimizzazione diretta della funzione (1.1) a causa soprattutto delle maggiori difficoltà operative che spesso impone a parità di efficienza nei risultati.

Può succedere comunque che le differenti caratteristiche numeriche degli algoritmi originati dai metodi risolutivi del sistema (1.3) risultino più aderenti al problema considerato e che abbiamo quindi maggiori possibilità di successo che non i metodi di minimizzazione diretta. Quale approccio seguire è una decisione che può scaturire solo da considerazioni contingenti che prescindono in buona parte dalle teorie dei metodi rimanendo basate sull'esame del particolare problema affrontato e sulle caratteristiche numeriche degli algoritmi effettivamente disponibili.

2. Due metodi lineari : Gauss-Seidel e Jacobi

Tra le molte tecniche per la soluzione di un sistema di equazioni lineari che potrebbero essere estese ai sistemi non-lineari ci limiteremo ad esaminarne due fra quelle più diffuse.

Metodo Gauss-Seidel

Sia dato il sistema lineare di k equazioni in k incognite:

$$\mathbf{Ax}=\mathbf{c} \quad \text{con} \quad \text{rango}(\mathbf{A}) = k \quad (2.1)$$

se gli elementi sulla diagonale della matrice \mathbf{A} sono tutti diversi da zero, le iterazioni procedono nel modo seguente :

Sia dato il valore x^p alla p -esima iterazione e sia pure dato il valore delle prime $(i-1)$ componenti di x^{p+1} alla $(p+1)$ -esima iterazione. Per ottenere il valore della i -esima componente di x^{p+1} bisogna risolvere rispetto ad x l'equazione lineare:

$$\sum_{j=1}^i a_{ij}x_j^{p+1} = a_{ii}x + \sum_{j=i+1}^k a_{ij}x_j^p = c_i \quad (2.2)$$

e porre $x_i^{p+1} = x$. Si passa poi ad un'altra equazione per ottenere la componente successiva x_{i+1}^{p+1} e questo si ripeterà fino a che non si siano determinate le k incognite e fino a quando le soluzioni non si siano ragionevolmente stabilizzate.

Una proficua modifica è quella di assumere x_i^{p+1} come dato dalla relazione:

$$x_i^{p+1} = x_i^p + \alpha(x - x_i^p) \quad (2.3)$$

dove x è la soluzione della (2.2) ed α è un parametro scelto in modo tale che:

$$a) \|Ax^{p+1} - c\| \leq \|Ax^p - c\|; \quad b) \lim_{p \rightarrow \infty} \|Ax^p - c\| = 0 \quad (2.4)$$

dove $\| \cdot \|$ indica una qualche norma definita su \mathbb{R}^k . Il principio espresso dalle (2.4) viene indicato come *Steepest-Descent* ed il suo rispetto per ogni valore dell'indice p costituisce un elemento essenziale della progettazione degli schemi iterativi.

Al fine di esprimere il processo Gauss-Seidel in forma compatta conviene scomporre la matrice A del sistema (2.1) nella forma

$$A = D - L - U \quad (2.5)$$

dove D è una matrice diagonale, L una matrice strettamente triangolare bassa ed U una matrice strettamente triangolare alta. L'avverbio strettamente indica che gli elementi sulla diagonale di L e di U sono degli zeri. Nell'ipotesi che gli elementi sulla diagonale di D siano tutti non nulli i passi (2.2) e (2.3) possono essere formulati nella più compatta espressione

$$\mathbf{x}^{p+1} = \mathbf{x}^p + \alpha(D - L)^{-1} \mathbf{r}^p \quad \text{con} \quad \mathbf{r}^p = A\mathbf{x}^p - \mathbf{c} \quad (2.6)$$

Metodo di Jacobi

Sia dato il valore di \mathbf{x}^p alla p -esima iterazione. La componente i -esima di \mathbf{x}^{p+1} alla $(p+1)$ -esima iterazione è determinata risolvendo rispetto ad x l'equazione lineare:

$$a_{ii}x + \sum_{\substack{j=i+1 \\ j \neq i}}^k a_{ij}x_j^p = c_i \quad (2.7)$$

e ponendo $x_i^{p+1} = x$. Anche in questa procedura è possibile utilizzare un parametro di guida che comporti la verifica delle (2.4). In questo caso l' i -esima componente alla iterazione $(p+1)$ sarà data da:

$$x_i^{p+1} = x_i^p + \alpha(x - x_i^p) \quad (2.8)$$

dove x è, analogamente all'altro schema, la soluzione della (2.7). Completata la determinazione delle k incognite si può procedere all'esame della differenza tra \mathbf{x}^p e \mathbf{x}^{p+1} per decidere se effettuare o meno un altro ciclo di iterazioni.

Utilizzando la scomposizione (2.5) la versione più compatta del metodo di Jacobi è

$$\mathbf{x}^{p+1} = \mathbf{x}^p + \alpha \mathbf{D}^{-1} \mathbf{r}^p \quad \text{con} \quad \mathbf{r}^p = \mathbf{A} \mathbf{x}^p - \mathbf{c} \quad (2.9)$$

Sia lo schema (2.6) che il (2.9) possono essere utilizzati nei sistemi di equazioni non lineari secondo due diverse impostazioni che assegnano ciascuna un preciso e differente significato alla locuzione "metodi lineari generalizzati" con cui questi tipi di schemi sono conosciuti.

3. Dualità dei metodi lineari generalizzati

Molti metodi per il calcolo delle regressioni non lineari, sia che aderiscano all'approccio della minimizzazione diretta della $S(\beta)$ in (1.1) che alla soluzione delle equazioni normali in (1.3) prevedono nel loro passo iterativo la soluzione di un sistema di equazioni lineari. Ad esempio il metodo Newton per la soluzione del sistema (1.3) adopera come incremento del valore corrente β^p la soluzione del sistema lineare

$$[H(\beta)]^{-1} (\beta^{p+1} - \beta^p) = q(\beta^p) \quad (3.1)$$

dove $H(\beta^p)$ è la matrice Hessiano costituita con le derivate parziali delle equazioni $\{q_i\}$ rispetto ai parametri $\{\beta_i\}$. Un primo modo di impiegare i metodi (2.6) e (2.9) è nella soluzione dei sistemi lineari di tipo (3.1) generati dai metodi non lineari. Si potrebbero ad esempio considerare combinazioni del tipo Newton-Seidel oppure Newton-Jacobi. Questa angolatura però coglierebbe i metodi lineari generalizzati nel loro ruolo tradizionale e sotto questo aspetto non presentano novità interessanti, anzi ci sono seri dubbi sulla loro efficienza soprattutto se messi a confronto con le tecniche di inversione della matrice del sistema (3.1).

L'estensione più promettente dei metodi iterativi lineari Seidel e Jacobi si raggiunge ipotizzando schemi misti non lineari/lineari. Ad esempio con il metodo Seidel si potrebbe costruire l'algoritmo:

$$\left\{ \begin{array}{l} 1. \text{ Dato } \beta^p = (\beta_1^p, \beta_2^p, \dots, \beta_{i-1}^p, \beta_i^p, \dots, \beta_k^p) \\ 2. \text{ Risolvi } q_i(\beta^p) = 0 \text{ rispetto a } \beta_i^+ \\ 3. \text{ Definisci } \beta_i^{p+1} = \beta_i^p + \alpha(\beta_i^+ - \beta_i^p) \\ 4. \text{ Se } i < k \text{ ritorna al punto 1} \end{array} \right. \quad (3.2)$$

Un simile schema sarebbe però valido se le equazioni ad una incognita che di volta in volta risultano al passo 2 avessero una soluzione unica in un qualche specifico dominio. Tale condizione è piuttosto difficile da assicurare a meno di ipotesi molto forti sulle funzioni coinvolte. Inoltre, anche se questa unicità di soluzione fosse garantita, non sempre sarebbe disponibile un algoritmo che converga alle singole soluzioni in un numero predefinito e comunque contenuto di iterazioni. E' possibile però applicare ogni volta il metodo di Newton (o altro equivalente) per ottenere delle approssimazioni alle soluzioni

delle singole equazioni. In questo caso il metodo di Newton assumerebbe la veste di componente secondaria rispetto all'impostazione primitivamente lineare del metodo. Definiamo quindi gli schemi

Algoritmo Seidel-Newton

$$\left\{ \begin{array}{l} 1. \text{ Dato } \beta^p = (\beta_1^{p+1}, \beta_2^p, \dots, \beta_{i-1}^p, \beta_i^p, \dots, \beta_k^p) \\ 2. \text{ Calcola } h_i(\beta^p) = [H(\beta^p)]^{-1} q_i(\beta^p) \\ 3. \text{ Definisci } \beta_i^{p+1} = \beta_i^p + \alpha h_i(\beta^p) \\ 4. \text{ Se } i < k \text{ ritorna al punto 1} \end{array} \right. \quad (3.3)$$

Algoritmo Jacobi-Newton

$$\left\{ \begin{array}{l} 1. \text{ Dato } \beta^p = (\beta_1^{p+1}, \beta_2^p, \dots, \beta_{i-1}^p, \beta_i^p, \dots, \beta_k^p) \\ 2. \text{ Calcola } h_i(\beta^p) = [H(\beta^p)]^{-1} q_i(\beta^p) \\ 3. \text{ Definisci } \beta_i^{p+1} = \beta_i^p + \alpha h_i(\beta^p) \\ 4. \text{ Se } i < k \text{ ritorna al punto 1} \end{array} \right. \quad (3.4)$$

La dualità di impiego dei metodi lineari generalizzati non è apprezzabile se si riferisce allo schema Jacobi-Newton. Infatti, esso è identico (come è facilmente provabile) all'algoritmo che si otterrebbe applicando il metodo Newton-Jacobi: cioè è indifferente quale metodo si utilizza come componente primaria e quale come secondaria. La dualità è invece piena nel caso del metodo Seidel-Newton dove la fase lineare delle iterazioni comporta elementi nuovi rispetto alla solita impostazione delle procedure di stima non lineare che ne suggeriscono una positiva attenzione nella generale metodologia di calcolo delle regressioni non lineari.

4. Applicazioni

Considerazioni sugli algoritmi

La sperimentazione delle procedure Seidel-Newton e Jacobi-Newton è stata affiancata dall'applicazione del metodo di Newton che prevede il seguente schema iterativo:

$$\beta^{p+1} = \beta^p + \alpha [H(\beta)]^{-1} q(\beta^p) \quad (4.1)$$

Le prove sono state organizzate in modo di mantenere, per quanto possibile, uniformità nella programmazione ed omogeneità di trattazione tali da rendere comparabili i metodi. In particolare si sono bloccate certe opzioni (numero massimo di iterazioni, test di convergenza, scritture intermedie, etc.) che vengono di solito variate di prova in prova.

Ecco le scelte fatte:

- 1) Funzione obiettivo: $\|q(\beta)\| = \sum_{i=1}^k |q_i(\beta)|$
- 2) Numero massimo di iterazioni: 50.
- 3) Criteri di stabilità numerica ovvero situazioni in cui la procedura deve essere interrotta:
 - a) Numero massimo di aumenti consecutivi ammessi nella funzione obiettivo: 5.
 - b) Numero massimo di aumenti nella funzione obiettivo: 15.

$$a) \frac{\|h\|}{[\gamma_1 + \|b\|]} \leq 0.000005; \text{ con } \|h\| = \sum_{i=1}^k |\beta_i^{p+1} - \beta_i^p|$$

- 4) Test di convergenza:
 - b) $\frac{\|q(\beta)\|}{k} \leq 0.0000005$

5) Derivate parziali. E' stato previsto il calcolo analitico sia delle derivate prime che delle derivate seconde della funzione di risposta del modello. Dato il ridotto numero di parametri che comparivano nei modelli di prova questa è sembrata la scelta più efficiente. In generale sono da preferire routine che forniscano approssimazioni numeriche ragionevoli delle derivate parziali.

6) Scelta di α . Il valore di a è stato calcolato ad ogni iterazione in base alla ottimizzazione parziale

$$\alpha_p = \underset{\alpha \in I(\alpha)}{\text{Minimo}} \left\{ q(\beta^p - \alpha d^p) \mid q(\beta^p - \alpha d^p) \in B \right\}, \quad d^p = [H(\beta)]^{-1} q(\beta^p)$$

$$I(\alpha) = \{0.01, 0.25, 0.50, 0.75, 1.0\}$$

Si è in pratica scelto il valore di a tra quelli in $I(\alpha)$ che produceva il maggior decremento nella funzione obiettivo fermo restando che il vettore dei parametri così determinati rimanesse nell'insieme dei valori ammissibili.

Le condizioni teoriche della convergenza e la dimostrazione dei teoremi che le incorporano sono discusse in modo esauriente nell'eccellente testo di Ortega e Rheinboldt (1970). Senza scendere troppo nei dettagli è possibile dire che le assunzioni fatte nel primo paragrafo e la tecnica di scelta del parametro a assicurano da un lato che gli schemi esaminati siano sempre well-defined cioè che da essi non verranno sempre prodotti valori ammissibili e che la sequenza con cui essi vengono generati converge sicuramente al punto di ottimo. Bisogna comunque considerare che nella maggioranza dei casi si potrà ottenere solo una aderenza incompleta alle assunzioni che garantiscono l'ottimalità. Questo porta in primo piano, rispetto alle condizioni teoriche di convergenza, le possibilità di convergenza sul valore teorico dei parametri e la robustezza delle procedure rispetto alle difformità dalle condizioni che assicurano l'ottimalità. Le applicazioni effettuate in questo lavoro, per come sono state progettate e realizzate, offrono concrete indicazioni sul funzionamento dei metodi lineari generalizzati.

Costruzione dei modelli di prova

La verifica empirica delle tecniche proposte nei paragrafi precedenti è stata condotta sui cinque modelli seguenti scelti fra quelli più ricorrenti nelle regressioni non lineari:

$$\begin{aligned}
A: y_i &= \beta_1 x_{i1} + \beta_2 e^{-\beta_3 x_{i2}} + e_t; & B: y_i &= \beta_1 e^{\beta_2 x_{i1}} + \beta_3 e^{\beta_4 x_{i2}} + e_t \\
C: y_i &= \beta_1 x_{i1} + \beta_2 (\beta_3)^{x_{i2}} + e_t; & D: y_i &= \beta_1 (\beta_2)^{x_{i1}} + \beta_3 (\beta_4)^{x_{i2}} + \beta_5 (\beta_6)^{x_{i3}} + e_t \\
E: y_i &= \beta_1 (x_i)^{\frac{\beta_2}{1-\beta_3}} + e_t
\end{aligned}$$

I valori teorici dei parametri sono riportati nella tabella 1.1; essi sono arbitrari, ma coerenti con l'ordine di grandezza dei regressori. L'ampiezza campionaria ed i valori dei regressori sono stati ottenuti trascrivendo i dati pubblicati nelle applicazioni dei modelli A-E. Nelle prime colonne della tabella 1.2 si riportano, ad ogni buon conto, le loro medie e varianze. I valori degli $\{e_t\}$ sono stati simulati con il metodo Monte Carlo da una distribuzione normale con media nulla e varianza riportata nell'ultima colonna della tabella.

Tabella_1.1: simulazione dei modelli di prova. Valori dei parametri.

| Modello | β_1 | β_2 | β_3 | β_4 | β_5 | β_6 | $\sigma^2(e)$ |
|---------|-----------|-----------|-----------|-----------|-----------|-----------|---------------|
| A | 3.00 | 2.00 | 1.00 | | | | 0.15 |
| B | 4.20 | -0.25 | 3.10 | -0.58 | | | 0.25 |
| C | 10.00 | -2.50 | 2.00 | | | | 0.05 |
| D | 3.18 | 2.07 | -4.15 | 3.27 | -2.88 | 1.11 | 0.75 |
| E | 10.00 | 0.78 | 0.03 | | | | 0.25 |

Tabella_1.2: simulazione dei modelli di prova. Valori dei regressori.

| Modello | X1 | | X2 | | X3 | | Ampiezza campionaria |
|---------|---------|------------|---------|------------|----------|------------|----------------------|
| | μ_1 | σ_1 | μ_2 | σ_2 | μ_3 | σ_3 | |
| A | 0.36 | 0.15 | 0.28 | 0.19 | | | 12 |
| B | 1.25 | 0.74 | 0.34 | 0.27 | | | 20 |
| C | 21.31 | 15.63 | 5.86 | 2.69 | | | 7 |
| D | 8.74 | 12.78 | 0.47 | 0.31 | 5.43 | 0.82 | 34 |
| E | 3.74 | 7.26 | -4.2602 | -0.9289 | -26.9184 | -9.2459 | 15 |

Come valore *good* della stima d'avvio dei parametri è stato scelto quello teorico e come valore iniziale *bad* si è scelto il valore teorico diviso per dieci.

5. Risultati e commenti.

Nelle tabelle 2.1-2.5 sono riassunti i risultati ottenuti nelle applicazioni dei processi Seidel-Newton, Jacobi-Newton e Newton. Per seguire il commento conviene avere in mente l'espressione analitica della funzione di risposta.

Nel modello A le condizioni di ottimo erano più fedelmente rispettate, soprattutto con valori di partenza *good* e questo ha comportato il successo di tutti i tre metodi. Un risultato confortante poiché è un riscontro empirico della effettiva validità delle condizioni teoriche espresse nel paragrafo 1. Con valori iniziali *bad* solo il processo Seidel-Newton è riuscito ad individuare il punto di ottimo. Il metodo Jacobi-Newton si è invece attestato su di un punto che non solo non era di ottimo, ma che si collocava anche al di fuori dell'insieme dei valori ammissibili. Il metodo di Newton rispetto ai valori *bad* ha reagito

" esplodendo" dopo cinque iterazioni.

Nel modello B erano assenti i termini puramente lineari, ma erano invece presenti due termini esponenziali di struttura simile. Modelli di questo tipo provocano quasi sempre delle inefficienze nei metodi del gradiente con associata inversione di matrice (tipo Newton, in pratica. Infatti, la multicollinearità accentuata dal doppio termine esponenziale si è rivelata, per la tecnica di Newton, un fattore destabilizzante sul sentiero della convergenza che si è avuta su un punto non di ottimo con valori iniziali *bad*, non si è avuta del tutto con valori iniziali *good*. Nello stesso modello, con valori iniziali *good* sia il metodo Seidel-Newton che lo Jacobi-Newton hanno avuto successo. Con valori iniziali *bad* il metodo Seidel-Newton ha mantenuto la sua validità laddove il metodo Jacobi-Newton, pur generando dei valori non lontani dal valore teorico dei parametri, ha comportato troppe oscillazioni nella funzione guida ed è stato interrotto.

Tabella 2.1: risultati per il modello A

| Met. | V.I. | β_1 | β_2 | β_3 | $\ S(\beta)\ $ | Iter. | (n-k)s(β) |
|------|------|-----------|-----------|-----------|----------------------|-------|-------------------|
| SN | B | 3.2225 | 2.1039 | 1.2153 | 0.0029 | 24 | 0.40 |
| | G | 3.2224 | 2.1040 | 1.2153 | 0.0019 | 24 | 0.41 |
| JN | B | 4.5655 | 5818.2800 | 64.8352 | 0.3×10^{-4} | 50 | 6.20 |
| | G | 3.2220 | 2.1042 | 1.2152 | 0.0020 | 46 | 0.41 |
| N | B | -3.5355 | 0.4952 | 8.9473 | 33.4540 | 5 | instabile |
| | G | 3.1963 | 2.0452 | 1.1315 | 0.1×10^{-5} | 4 | 0.40 |

Tabella 2.2: risultati per il modello B

| Met. | V.I. | β_1 | β_2 | β_3 | β_4 | $\ S(\beta)\ $ | Iter. | (n-k)s(β) |
|------|------|-----------|-----------|-----------|-----------|----------------------|-------|-------------------|
| SN | B | 3.7728 | -0.2870 | 3.0719 | -0.3653 | 3.6702 | 50 | 4.13 |
| | G | 4.1074 | -0.2258 | 3.0377 | -0.5659 | 5.0946 | 22 | 1.78 |
| JN | B | 3.7510 | -0.3921 | 4.1640 | -0.3310 | 229.2320 | 6 | instabile |
| | G | 4.0370 | -0.2304 | 2.9023 | -0.5712 | 1.1895 | 10 | 6.14 |
| N | B | 5.4616 | -97.8831 | 0.0078 | -27.6063 | 0.0001 | 15 | 384.29 |
| | G | 27.5476 | 5.5681 | -48.7458 | -1.1741 | 0.1×10^{37} | 9 | instabile |

Tabella 2.3: risultati per il modello C

| Met. | V.I. | β_1 | β_2 | β_3 | $\ S(\beta)\ $ | Iter. | (n-k)s(β) |
|------|------|-----------|-----------|-----------------------|----------------|-------|-------------------|
| SN | B | 9.3032 | 139.1834 | 0.1514 | 4.6028 | 15 | 911.16 |
| | G | 10 | -2.0015 | 2.0002 | 1.8632 | 1 | 0.01 |
| JN | B | 9.529 | -17.3964 | -0.2893 | 471.0382 | 1 | instabile |
| | G | 10 | -2.0008 | 2.0004 | 2.0276 | 1 | 0.01 |
| N | B | 9.6705 | -0.0014 | -0.9×10^{-9} | 0.0138 | 4 | 1214.71 |
| | G | 10 | -2.0017 | 2.0000 | 2.0145 | 1 | 0.00 |

Nel modello C erano messe in risalto, oltre alla scontata presenza di multicollinearità presente in ogni modello non lineare, diversi altri elementi caratteristici del calcolo delle regressioni non lineari. La varianza dei residui piuttosto piccola che è stata adottata nella simulazione dei residui ha reso il modello virtualmente deterministico. Questo è in genere un fattore positivo che favorisce la procedura di calcolo. Tuttavia, la ridotta numerosità del campione ha limitato di molto i vantaggi. In questi casi la componente non lineare del modello C può essere più facilmente definita attraverso combinazioni di valori,

Tabella 2.4: risultati per il modello D

| Metodo | V.I. | β_1 | β_2 | β_3 | β_4 | β_5 | β_6 | $ Q(\beta) $ | Iter. | $(n-k)\sigma^2$ |
|--------|------|----------------------|-----------|----------------------|-----------|------------|-----------|----------------------|-------|---------------------|
| SN | B | -26.6775 | 0.2189 | 0.1x10 ¹⁴ | -2.6334 | -1357.7100 | 55.5488 | 0.1x10 ⁷ | 12 | instabile |
| | G | 3.1745 | 2.0709 | -4.1499 | 3.2700 | -2.6521 | 0.6773 | 2100.5436 | 20 | 14.1783 |
| JN | B | -0.1x10 ⁶ | -2.4450 | -657.2528 | -3.0252 | 3957.2375 | 6.5833 | 0.8x10 ⁸ | 10 | instabile |
| | G | 3.0303 | 2.0505 | -4.2087 | 3.2829 | -28.7428 | 3.6487 | 0.7x10 ⁷ | 10 | instabile |
| N | B | 0.3434 | -9.4513 | 0.5180 | -12.5064 | -1.7205 | 345.3248 | 0.0001 | 21 | 0.2x10 ⁸ |
| | G | 4.5332 | -1.0557 | -4.2602 | -0.9289 | -26.9184 | -9.2459 | 0.2x10 ¹¹ | 12 | instabile |

Tabella 2.5: risultati per il modello E

| Metodo | V.I. | β_1 | β_2 | β_3 | $ Q(\beta) $ | Iter. | $(n-k)\sigma^2$ |
|--------|------|---------------------|-----------|-----------|----------------------|-------|-----------------|
| SN | B | 20.2532 | 0.2683 | -0.0381 | 5.3426 | 50 | 88.6129 |
| | G | 8.3677 | 0.7864 | 0.0491 | 1.8362 | 14 | 138.1765 |
| JN | B | 0.2x10 ⁵ | -233.2661 | -329.5614 | 135.0237 | 22 | instabile |
| | G | 17.2297 | 7.0397 | -17.1394 | 91.3126 | 14 | instabile |
| N | B | -9.0166 | 123.8684 | 8.5766 | 0.7x10 ⁻⁶ | 18 | 1333.8635 |
| | G | 11.3121 | 1.0104 | -5933 | 0.0059 | 19 | 69.5046 |

ciascuna associata ad una diversa soluzione delle equazioni normali, ma di cui soltanto una è relativa allo stimatore consistente. Ne consegue che nei modelli non-lineari l'ampiezza campionaria ha un ruolo ancora più determinante che nei modelli lineari poiché influenza non solo l'affidabilità delle stime, ma rende anche più complessa e disagiata la determinazione delle stime stesse. Nel modello C, con valori iniziali *good*, tutti i processi hanno raggiunto il punto di minimo assoluto con pochissime iterazioni; con valori iniziali *bad* ogni processo è arrivato ad un punto diverso in dipendenza della propria struttura numerica. E' interessante il fatto che le stime associate al parametro lineare siano rimaste stabili indipendentemente dal metodo e dal valore iniziale adottato laddove i parametri non lineari hanno presentato ampie oscillazioni in tutti i processi.

Le considerazioni fatte sui primi tre modelli sono da estendere anche ai modelli D ed E notando in più che la frequenza degli insuccessi dei metodi lineari generalizzati aumenta con il crescere della varianza dei residui utilizzata per la simulazione e con l'aumentare della "non linearità" delle funzioni di risposta. Infatti, man mano che $s^2(e)$ sale, i modelli risultano virtualmente stocastici. Il cattivo adattamento può introdurre forti elementi di disturbo alla convergenza come punti critici della somma dei residui quadrati la cui presenza è dovuta solo a fluttuazioni campionarie e non hanno nulla a che vedere con il punto critico definito dal valore vero dei parametri. La seconda osservazione è da ritenere come elemento indicativo di reticenza nell'utilizzo di schemi in cui la validità del passo iterativo dipende dal più o meno scarso condizionamento di una qualche matrice da invertire senza che si possa controllarne in qualche modo il rango.

6. Conclusioni

Le indicazioni ottenute dalle prove effettuate possono essere riassunte nei seguenti punti: 1) il principio di utilizzare le informazioni di stima non appena esse si rendono disponibili -incorporato nel metodo Seidel-Newton- si è rivelato un fattore importante di stabilità numerica ed un contributo alla convergenza del processo. Questo rende il metodo Seidel-Newton effettivamente alternativo a tutti gli altri esistenti. La relativa lentezza in termini di numero di iterazioni è compensata dal minor tempo complessivo impiegato nonché dal minor spazio di memoria occupata.

2) Il metodo Jacobi-Newton così come è stato progettato non è efficiente, soprattutto con i modelli che includono parametri lineari. C'è però da aggiungere che la possibile applicazione nella veste Newton-Jacobi, lo rendono apprezzabile, opportunamente modificato, quando si affrontino problemi con moltissimi parametri.

3) il metodo di Newton nella versione 4.1 è chiaramente inadeguato per i problemi a forte non-linearità. La modifica più opportuna è quella suggerita da Marquardt di inserire un parametro sulla diagonale della matrice Hessiano che dovrebbe attenuare i problemi di multicollinearità e fornire una più adeguata caratterizzazione dell'intero processo.

4) I metodi lineari generalizzati come pure molti metodi suggeriti per la soluzione dei sistemi di equazioni non lineari sono di difficile gestione.

L'aspetto analitico essenziale è la matrice Hessiano: se questa viene introdotta analiticamente è quasi sempre in forma assai complessa e traducibile con difficoltà in linguaggio di programmazione; se invece viene approssimata con il calcolo numerico si rischia di introdurre ulteriori elementi turbativi della convergenza.

In definitiva, le prove gettano più ombre che luci sui metodi lineari generalizzati, tuttavia, nella particolare processo Seidel-Newton possono fornire un ulteriore tentativo da effettuare quando i metodi usuali non hanno funzionato in modo soddisfacente.

Bibliografia

- 1) Draper N.R., Smith H. (1970): Applied Regression Analysis. John Wiley & Sons, New York.
- 2) Foutz V. R. (1977): Existence of a Unique Consistent Solution to the Likelihood Equation. Journal of the American Statistical Association. 72.
- 3) Gallant A.R. (1975): Non-linear Regression. The American Statistician. 29.
- 4) Goldfeld S.M., Quandt R.E., Trotter H.F. (1966): Maximization by Quadratic Hill-Climbing. Econometrica. 34
- 5) Hartley H.O. (1961): The Modified Gauss-Newton Method for the Fitting of Non-linear Regression Functions by Least Squares. Technometrics. 3.
- 6) Jenrich I.R. (1969): Asymptotic Properties of Non-linear Least Squares Estimators. Annals of Mathematical Statistics. 40.
- 7) Kale B.K. (1962): On the Solution of Likelihood Equations by Iteration Process the Multiparameter Case. Biometrika. 19.
- 8) Kleijnen J.P. (1974): Statistical Techniques in Simulation. Part I. Marcell Dekker Inc. New York.
- 9) Malinvaud E. (1969): Metodi statistici per l'econometria. UTET, Torino
- 10) Malinvaud E. (1970): The Consistency of Non-linear Regressions. Annals of Mathematical Statistics. 41.
- 11) Marquardt D.W. (1963): An Algorithm for Least Squares Estimation of Non-linear Parameters. SIAM Journal. 2.
- 12) Tarsitano A. (1979). Alcune considerazioni sulle regressioni non lineari. Dipartimento di Economia politica. Università della Calabria.
- 12) Ortega J .M., Rheinboldt (1970): Iterative Solution of Non-linear Equation in Several Variables. Academic Press. New York.