

Alcune considerazioni sulle regressioni non lineari(*)

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

Riassunto.

La stima dei parametri nei modelli non lineari richiede tecniche iterative in cui da una data stima iniziale si procede per miglioramenti successivi fino alla stima finale. Talvolta questo percorso ideale è interrotto da overflow numerici. In tali casi si può ricorrere alla procedura del simplesso di Nelder e Mead. Lo scopo di questo lavoro è di introdurre tale tecnica nella stima dei parametri del modello di regressione non lineare.

Keywords: regressione non lineare, metodo del simplesso

() Dipartimento di Economia politica - 1979.*

1. Introduzione

Al fine di meglio chiarire il senso del lavoro è opportuno rivedere i concetti di base delle regressioni non-lineari e richiamare quindi il modello generale di riferimento :

$$\mathbf{y} = f(\mathbf{X}; \boldsymbol{\beta}) + \mathbf{e} \quad (1)$$

dove \mathbf{X} rappresenti il vettore delle m variabili indipendenti o esogene, \mathbf{y} la variabile dipendente o endogena e $\boldsymbol{\beta}$ sia il vettore incognito dei k parametri del modello nella popolazione dei valori. Infine, si consideri \mathbf{e} un errore casuale la cui distribuzione sia incognita, ma che abbia media nulla e varianza finita. L'espressione $f(\mathbf{X}, \boldsymbol{\beta})$ esprime il legame che regola l'associazione fra le \mathbf{X} e la \mathbf{y} attraverso un'opportuna combinazione dei parametri $\boldsymbol{\beta}$. In altre parole $f(\mathbf{X}, \boldsymbol{\beta})$ è il modello che sintetizza la teoria del fenomeno e con cui si studiano i dati (\mathbf{y}, \mathbf{X}) . Una volta che si siano determinati gli elementi incogniti della (1), la relazione che si determina può essere adoperata per:

- 1) Interpolazione. Determinazione del valore della endogena che corrisponde ad un valore prefissato della esogena per ognuno dei valori osservati delle \mathbf{X} . Lo scopo potrebbe essere di ricostruire valori mancanti o di sostituire valori anomali della dipendente.
- 2) Estrapolazione. Determinazione del valore della variabile endogena che corrisponde ad un valore prefissato delle variabili esogene non effettivamente osservato. Lo scopo è di proiettare la relazione oltre i dati osservati.
- 3) Controllo. Determinazione del valore delle variabili esogene idoneo a determinare un fissato livello della endogena

Il termine di errore \mathbf{e} include sia gli errori di misurazione e di rilevazione sempre riscontrabili nei dati reali, che errori stocastici dovuti alla irriproducibilità matematica dei fenomeni reali.

Al modello (1) si associa la qualifica non-lineare dei parametri. Questo vuole dire che le derivate parziali (almeno una) rispetto ai parametri della $f(\mathbf{X}, \boldsymbol{\beta})$ dipendono ancora da uno o più elementi di $\boldsymbol{\beta}$. Ancora, nelle regressioni lineari di solito il numero dei parametri incogniti coincide (a parte il termine costante) con il numero di regressori. Nelle regressioni non lineari k è regolarmente diverso da m .

La problematica collegata a questo tipo di modelli appare in tutta la sua complessità quando si consideri la procedura per ottenere le stime ai minimi quadrati dei parametri della funzione di regressione. Nei modelli lineari la somma dei residui quadrati genera sempre delle curve di livello ellittiche, di forma simile e con comune centro nel punto coincidente con la stima ai minimi quadrati dei parametri. Inoltre la funzione da minimizzare è globalmente convessa e possiede un unico punto di minimo.

Nei modelli di regressione non-lineari, la somma dei residui quadrati:

$$S(\boldsymbol{\beta}) = \sum_{i=1}^n [y_i - f(\mathbf{X}_i; \boldsymbol{\beta})]^2 \quad (2)$$

può generare curve di livello di qualsiasi forma; la sua convessità non è garantita e possono coesistere, insieme al minimo globale, altri minimi locali o addirittura un *continuum* di

minimilocali. Comunque, quello che di solito succede è che in prossimità di un punto di minimo le curve di livello tendono a divenire ellittiche, la convessità della (2) si accentua ed è dunque possibile la stima non-lineare, qualora si disponga di un punto di partenza che ricada nella zona di convessità esistente intorno al minimo globale ed uno schema iterativo che sia in grado di raggiungerlo.

2. Procedure di calcolo delle regressioni non lineari

La stima dei parametri nei modelli non-lineari può essere ancorata a due idee-guida: una espressa da Gauss che già nel 1809 suggeriva di trattare i modelli non-lineari con una successione convergente di approssimazioni lineari; ed un'altra escogitata da Cauchy nel 1847 che sottolinea il fatto che dato un punto qualsiasi nello spazio delle incognite, la funzione da minimizzare decresce più rapidamente nella direzione del negativo del gradiente valutato in quel punto.

I due principi, possono essere ricondotti alla generale formula iterativa :

$$b^{q+1} = b^q - \gamma [R^q + \lambda I]^{-1} h^q \quad (3)$$

dove i b^q sono le stime dei parametri incogniti alla iterazione q-esima; γ e λ sono due numeri scalari (in particolare $\lambda > 0$) che determinano ad ogni iterazione, la lunghezza del passo di correzione. Il vettore h^q è il gradiente della funzione (2) valutato nel punto b^q . R^q è una matrice reale, simmetrica, non-negativa definita la cui scelta determina il particolare algoritmo (traiettoria delle approssimazioni successive) che si utilizza.

La parte più importante, nello schema (3), è riservata alle derivate parziali prime rispetto ai parametri: oltre a comparire nel gradiente h , in molti algoritmi (ad esempio i cosiddetti metodi Gauss-Newton), contribuiscono alla definizione della matrice R^q . In altre parole sono le derivate a determinare in modo essenziale l'incremento da dare alla stima corrente b^q per ottenerne una "migliore" b^{q+1} .

Il perchè si sottolinea il ruolo delle derivate non è semplice da spiegare se non con un linguaggio impreciso e povero di ancoraggi immediatamente accessibili. In molti casi la $f(X, \beta)$ ha un forte grado di non-linearità (l'espressione non è accurata, ma si allude, ad esempio alla presenza di termini esponenziali molto imbottiti di parametri) che si concreta frequentemente in correzioni h^q che pur includendo valori plausibili rispetto alla minimizzazione della (2) produce però termini (3) che danno luogo ad *overflow* numerici nel computer: ovvero nell'espletamento dei calcoli si generano numeri più grandi di quello che la macchina è in grado di ritenere correttamente. Ad esempio, nel termine β^x con valori della x intorno a 100 e un valore della stima $b=2.7$ un leggero spostamento oltre il 2.7 blocca subitaneamente la procedura di calcolo perché le parole del computer attualmente in uso non consentono di andare oltre e^{99} (dove e è il numero di Eulero). Subentra in pratica una nuova "Legge dei Grandi Numeri" che distrugge nella fase applicativa tanti algoritmi pure costruiti sulle migliori premesse formali.

Non ci sono trucchi o artifici che permettano di superare l'ostacolo dato agli *overflow* e contemporaneamente mantenere un sentiero iterativo plausibile percorribile da un algoritmo efficiente. Non basta ad esempio estendere la parola di memoria perchè la nuova legge dei grandi numeri è cumulativa: dopo qualche iterazione ci si troverà di fron-

te al problema iniziale. Non basta neanche ridurre la scala delle variabili incluse nel modello: il problema degli *overflow* riguarda il tipo di funzione e non le variabili. Questi tentativi potrebbero funzionare quando si è in grado, a priori, di localizzare il punto di minimo globale e di essere sicuri che intorno a quel punto la funzione (2) presenti una *mild-nonlinearity*. L'unica scappatoia realistica è di ricorrere a tecniche alternative che evitino l'uso delle derivate e di conseguenza limitino il rischio di esondazioni numeriche, senonaltro perchè le componenti funzionali complesse intervengono nei calcoli con frequenza notevolmente minore.

3. La procedura del semplice nelle regressioni non-lineari.

Fra i metodi di minimizzazione che per la loro struttura semplificata risultano come valida alternativa a metodi più solidi teoricamente, ma poco pratici, un posto di primo piano merita l'algoritmo proposto da Nelder e Mead (1965) particolarmente valido per modelli non-lineari che prevedano pochi parametri (5 al massimo). Si tratta di un metodo che utilizza solo la (2) in successive valutazioni secondo uno schema estremamente rarefatto: se b_1, b_2, b_3 sono punti dello spazio parametrico apparentemente ad R^k tali che $S(b_1) > S(b_2)$ e $S(b_1) > S(b_3)$ si esplora lo spazio parametrico in alcuni punti prefissati lungo il piano che congiunge b_1 al punto di mezzo di quello che congiunge b_2 e b_3 . Lo scopo è di determinare un punto b_4 tale che $S(b_4) < S(b_1)$. Una volta determinato un punto b_4 le iterazioni proseguono ragionando allo stesso modo sulla tema b_4, b_2, b_3 . In pratica, ad ogni passo, l'algoritmo costruisce un semplice ($k+1$ punti nello spazio a k dimensioni) il cui orientamento ed il cui volume vengono modificati ad ogni iterazione. Nel caso in cui non fosse possibile determinare il punto b_4 il semplice è contratto intorno a b_2 o b_3 secondo che $S(b_2) < S(b_3)$ o viceversa. La ricerca del minimo avviene perciò all'interno di una griglia più o meno sparsa di punti e non seguendo una superficie continua di valori.

L'algoritmo si interrompe quando il valore della somma dei minimi quadrati nei vertici del semplice è abbastanza uniforme in base ad un qualche criterio prefissato: ad esempio la varianza della (2) nei vertici del semplice. Esistono poi degli artifici che consentono di superare eventuali minimi locali anche se questo non è forse sufficiente a far dimenticare l'opportunità del metodo: ogni nuovo semplice cancella quello precedente cosicché la determinazione del minimo è legata ai soli risultati correnti e non è perciò connessa ad uno schema iterativo univocamente pre determinabile fin dalla prima iterazione

La versione più nota dell'algoritmo del semplice è quella fornita da R. O'Neill (1971) che è stata provata da diversi autori (si vedano i riferimenti bibliografici). In questo lavoro la tecnica è stata usata per stimare i parametri dei due modelli di regressione non-lineari :

$$\text{Modello logistico } y = \frac{\beta_1}{[1 + \beta_2 e^{-\beta_3 x}]^{\beta_4}} \quad (4)$$

$$\text{Legge di Mitscherlich sui rendimenti decrescenti: } y = \beta_1 + \beta_2 x^{\beta_3} \quad (5)$$

applicati alla serie dei tassi istantanei di mortalità secondo la funzione obbiettivo:

$$S(\mathbf{b}) = \sum_{i=1}^n w_i [y_i - f(\mathbf{X}, \mathbf{b})]^2 \quad (6)$$

dove gli w_i sono dei pesi, noti a priori, che orientano la stima \mathbf{b} determinando un adattamento selettivo del modello ai valori osservati.

I risultati sono esposti nelle Tabella 1 e Tabella 2 e sono da considerarsi più che soddisfacenti tenendo conto che sugli stessi dati e modelli sono rovinosamente caduti la migliore delle tecniche Gauss-Newton (illMarquardt's Algorithm) ed il noto Rank-Two di Davidon-Fletcher-Powell.

Tabella 1: risultati della procedura di stime nel modello logistico

Serie	S(b)	M (1)	b1	b2	b3	b4
1	92.4311	173'018	4.5228	36.5784	0.0389	0.7679
2	108.6418	61'056	5.2313	36.4503	0.0383	3.9986
3	56.7800	318	4.2970	44.8942	0.0400	3.4165
4	62.4442	612'233	3.5396	47.2099	0.0428	3.8504
5	39.6229	330	4.6497	39.0757	0.0383	3.4060
6	91.8224	233'026	14.2237	38.2337	0.0321	3.5101
7	52.3052	321	4.9448	46.2482	0.0429	3.6529
8	77.4190	318	4.6213	51.3959	0.0429	4.0656
9	14.7146	263'072	4.3130	77.5683	0.0393	2.1749
10	32.0222	213'049	2.3703	40.1474	0.0412	3.1482

(1) indica il numero di volte che è stata richiamata la routine che calcola S(b). Numeri elevati di M non devono sorprendere se si tiene conto che in presenza di sospettiminimi locali l' algoritmo procede ad una gragliatura esplorativa dello spazio parametrico dopo aver aumentato M di 10000.

Tabella 2: risultati delle stime nel modelli Mitscherlich

Serie	S(b)	M	b1	b2	b3
1	92.8423	310	-0.0590	0.0012	1.0647
2	170.0640	143	-0.0772	0.0014	1.0619
3	57.2934	340	-0.0651	0.0016	1.0619
4	62.3068	11	-0.0858	0.0015	1.0619
5	40.7581	331	-0.0517	0.0015	1.0621
6	97.4446	163	-0.0556	0.0014	1.0621
7	51.9624	334	-0.0893	0.0016	1.0620
8	84.5769	11	-0.1194	0.0016	1.0619
9	32.7741	183	0.2525	0.0007	1.0619
10	64.8503	183	0.2242	0.0006	1.0619

Bibliografia

- Benyon P.R.(1976). Remark AS R15 *Applied Statistics*, 25, 97
- Chambers J.M. Ertel J.E. (1974). Remark AS R11. *Applied Statistics*, 23, 250-251,
- Fletcher R., Powell M.J.D. A(1963). rapidly convergent descent method for minimization.
The Computer Journal, .6,163-168
- Hill I.D.(1978). Remark AS R28 *Applied Statistics*, 27, 380-382.
- Marquardt D. W. (1963). An algorithm for least square estimation of non linear parameters.
J.Soc.Indust.Appl.Math. 11, pp.431-444.
- Nelder J. Mead R. (1965). A simplex method for function minimization. *The Computer Journal*, 7, 308-313.
- O'Neill R. (1971). Function minimization using a simplex procedure. Algorithm AS 47
Applied Statistics, 20, 338-345