

LE RETI NEURALI PER LA SCELTA DELLA FINESTRA NELLA STIMA NONPARAMETRICA DI DERIVATE

F. Giordano¹, M.L. Parrella¹

¹ Dipartimento di Scienze Economiche e Statistiche
Università degli Studi di Salerno
(e-mail: giordano@unisa.it, mparrella@unisa.it)

ABSTRACT: In questo lavoro viene preso in considerazione il problema della scelta della finestra di smoothing nell'ambito della stima di derivate di funzioni mediante i polinomi locali. Il contesto analizzato è quello delle serie storiche derivanti da processi strettamente stazionari, di cui si considera la stima della derivata prima della media condizionale. Mediante un esempio su dati simulati, viene evidenziata la crucialità della scelta della finestra di smoothing sul risultato ottenuto dalla stima nonparametrica. Viene, quindi, proposto l'utilizzo di una particolare tecnica di selezione di tale parametro basata sull'uso delle reti neurali.

KEYWORDS: scelta della finestra di smoothing, dati dipendenti, polinomi locali, reti neurali.

1 Introduzione

Sia $\{X_t; t = 1, \dots, n\}$ una realizzazione di lunghezza n da un processo stocastico stazionario a valori reali, $\{X_t; t \in \mathbb{N}\}$. Si consideri il seguente modello autoregressivo

$$X_t = m(X_{t-1}) + \varepsilon_t \quad (1)$$

dove $\{\varepsilon_t\}$ sono errori *i.i.d.* con media 0 e varianza $\sigma^2 \in (0, +\infty)$. Si supponga che la funzione $m(x)$ sia dotata di derivata tera continua. Il modello (1) è utile per modellare strutture nonlineari comunemente incontrate nelle serie storiche finanziarie ed econometriche. La media condizionale $m(x)$ e le sue derivate possono essere stimate in modo nonparametrico mediante i polinomi locali. Questi stimatori godono di buone proprietà teoriche, ma sono soggetti alla corretta scelta di un parametro di smoothing, la finestra o *bandwidth* della funzione kernel, che di fatto può comprometterne sostanzialmente la performance. In questo lavoro viene evidenziata empiricamente la difficoltà della scelta di tale parametro, presentando i risultati di stima ottenuti su alcune serie simulate. Viene quindi proposto l'utilizzo di un metodo plug-in per la scelta della finestra alternativo a quelli comunemente usati, basato sull'utilizzo delle reti neurali. Tale metodo permette di superare alcuni limiti che generalmente si incontrano con i metodi di scelta tradizionali.

2 Scelta della finestra di smoothing

Si consideri lo stimatore dei polinomi locali di grado p della funzione $m'(x)$, per $x \in \mathbb{R}$. Per semplicità consideriamo $p = 2$, solitamente usato per la stima della derivata prima.

Tale stimatore si ottiene risolvendo il seguente problema di minimizzazione

$$\arg \min_{\beta_0, \beta_1, \beta_2} \sum_{t=2}^n \left(X_t - \sum_{j=0}^2 \beta_j(x) (X_{t-1} - x)^j \right)^2 K \left(\frac{X_{t-1} - x}{h_n} \right)$$

rispetto a $\beta_0(x)$, $\beta_1(x)$ e $\beta_2(x)$. Si noti che $\hat{m}_h(x) \equiv \hat{\beta}_0(x)$, $\hat{m}'_h(x) \equiv \hat{\beta}_1(x)$ e $m''_h(x) \equiv 2\hat{\beta}_2(x)$. La funzione kernel $K(u)$ è una funzione di densità definita nel compatto $[-1, 1]$ ed è influenzata dal parametro bandwidth h_n . È chiaro che h_n rappresenta un parametro di scala per la funzione peso $K(\cdot)$ e determina l'ammontare di punti sostanzialmente considerati nella somma e quindi il grado di *smoothness* della funzione stimata.

Il metodo proposto per la scelta di h_n si basa su alcuni risultati presentati in Giordano e Parrella (2006). Consideriamo la seguente quantità

$$\widetilde{MISE}(h) = \int \{ Bias^2 [\hat{m}'_h(x)] + Var [\hat{m}'_h(x)] \} w(x; \tau) f_X(x) dx, \quad (2)$$

in cui $\hat{m}'_h(x)$ è la stima locale lineare della funzione $m'(x)$, $f_X(x)$ rappresenta la densità di X_t , mentre la funzione peso $w(x; \tau)$ è pari alla densità della distribuzione normale con media zero e varianza τ^2 pari alla varianza di X_t . La (2) rappresenta una versione modificata dell'errore quadratico medio integrato. Masry e Fan (1997) indicano distorsione e varianza asintotica dello stimatore $\hat{m}'_h(x)$, grazie ai quali è semplice ricavare la bandwidth asintoticamente ottimale minimizzando la (2):

$$h_{AMISE} = \left(\frac{9\sigma^2 V_1}{2B_1^2 R(m''')} \right)^{1/7} n^{-1/7}, \quad (3)$$

dove $R(m''') = \int [m'''(x)]^2 w(x; \tau) f_X(x) dx$. Le quantità B_1 e V_1 dipendono dal kernel e sono quindi note, mentre le quantità σ^2 e $R(m''')$ non sono note e vanno quindi stimate. Data l'ergodicità del processo, proponiamo i seguenti due stimatori

$$\hat{R}_{NN}(m''') = \frac{1}{n-1} \sum_{t=2}^n [g'''(X_{t-1}; \hat{\eta})]^2 w(X_{t-1}; \hat{\tau}); \quad \hat{\sigma}_{NN}^2 = \frac{1}{n-1} \sum_{t=2}^n \hat{\varepsilon}_t^2 \quad (4)$$

dove $\hat{\tau} = n^{-1} \sum (X_t - \mu)^2$, $\hat{\varepsilon}_t = X_t - g(X_{t-1}; \hat{\eta})$, e $g(x; \hat{\eta})$ e $g'''(x; \hat{\eta})$ rappresentano rispettivamente la stima di $m(x)$ e della sua derivata terza $m'''(x)$ ottenute mediante le reti neurali. Indichiamo con \hat{h}_{NN} lo stimatore della finestra proposto, ottenuto sostituendo nella (3) le stime ottenute in (4).

3 Applicazione a dati simulati

Consideriamo i seguenti due modelli autoregressivi AR(1) ed EXPAR(1)

$$(a) \quad X_t = 0.8X_{t-1} + \varepsilon_t \quad (b) \quad X_t = (0.6 + 0.6e^{-X_{t-1}^2})X_{t-1} + \varepsilon_t.$$

È evidente che la media condizionale del modello AR(1) è lineare, mentre la media condizionale del modello EXPAR(1) è pari a $m(x) = (0.6 + 1.5e^{-x^2})x$, e quindi non-lineare. La figura 1 mostra l'andamento della funzione di media condizionale (linea

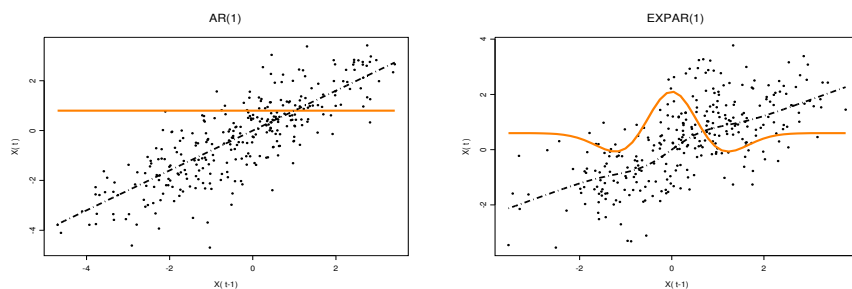


Figura 1. Scatter plot di due serie simulate dai modelli $AR(1)$ ed $EXPAR(1)$, insieme alle rispettive medie condizionali (---) e derivate prime delle medie condizionali(—).

tratteggiata) e della rispettiva derivata prima (linea continua) per i due modelli considerati, insieme allo scatter plot di una serie simulata da tali modelli, di lunghezza $n = 300$. Come si può notare, la non linearità del modello (b) appare evidente dall'andamento della derivata prima della media condizionale, più che dall'andamento della media condizionale stessa. La figura 2 riporta i grafici con i risultati di un esperimento condotto sulle due serie simulate. Le linee tratteggiate rappresentano le stime della derivata prima per i due modelli considerati, ottenute mediante i polinomi locali di grado $p = 2$, per valori crescenti del parametro bandwidth $h_n = (1, 2, 6)$. Le linee continue evidenziano l'andamento della funzione vera. I grafici a sinistra riguardano il modello lineare $AR(1)$, mentre i grafici a destra riguardano il modello non lineare $EXPAR(1)$. Appare chiaro che il valore di h_n influenza in modo cruciale i risultati della stima. Per il modello $AR(1)$, valori piccoli di h_n determinano una funzione stimata che appare molto irregolare e quindi lontana dalla funzione vera (costante), mentre valori grandi della finestra danno una buona stima della derivata prima. Per il modello $EXPAR(1)$ accade il contrario, per cui valori grandi di h_n determinano una funzione stimata che appare (erroneamente) costante, mentre valori piccoli della finestra forniscono una funzione stimata più irregolare ma vicina alla funzione vera (a parte alcuni inevitabili effetti di "bordo"). Questo dimostra che l'ampiezza della finestra del kernel non dipende soltanto dalla lunghezza della serie ma anche (ovviamente) da altri fattori non noti, quali appunto il funzionale $R(m''')$. Ne deriva che la scelta di tale parametro non è affatto semplice, sebbene sia fondamentale ai fini del corretto utilizzo degli stimatori dei polinomi locali. Si tenga presente che spesso la scelta di h_n viene fatta "a occhio" nelle pratiche applicazioni, e questo perchè i metodi tradizionali per la scelta automatica di tale parametro sono tutt'altro che semplici da implementare e/o tutt'altro che efficaci su campioni finiti.

La tabella nella pagina seguente riporta i risultati di uno studio di simulazioni effettuato al fine di analizzare empiricamente il funzionamento del metodo di scelta della finestra qui proposto. Accanto al valore vero per la finestra ottimale, indicato con h_{AMISE} , vengono riportati i quartili delle stime di \hat{h}_{NN} ottenute su 200 serie simulate

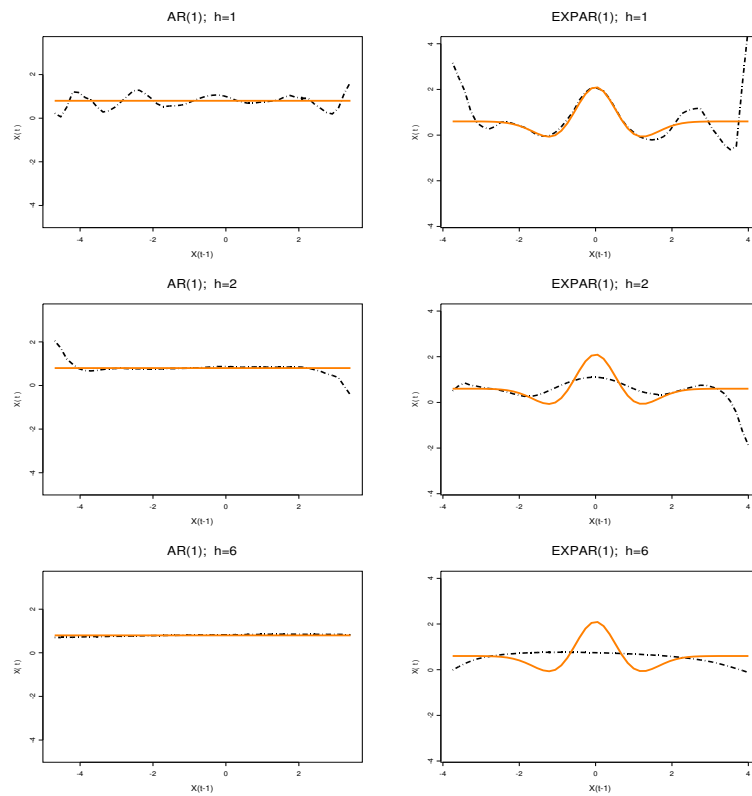


Figura 2. Stime della derivata prima (---) per il modello AR e per il modello EXPAR, ottenute mediante i polinomi locali di grado $p = 2$, per valori crescenti del parametro bandwidth h (dall'alto in basso). La linea continua evidenzia l'andamento della funzione vera.

da ciascun modello.

	h_{AMISE}	\hat{h}_{NN} con $n = 300$		
AR(1)	∞	$Q_1 = 2.708$	$Q_2 = 3.895$	$Q_3 = 9.827$
EXPAR(1)	0.642	$Q_1 = 0.535$	$Q_2 = 0.627$	$Q_3 = 0.787$

Bibliografia

- GIORDANO, F., & PARRELLA, M.L. 2006. Local Polynomial vs Neural Networks: some empirical evidences. *Atti del Convegno Naz. Ricerche sulle Serie Temporalì, Roma*.
- MASRY, E., & FAN, J. 1997. Local polynomial estimation of regression functions for mixing processes. *Scandinavian Journal of Statistics*, **24**, 165–179.