LA COMBINAZIONE DI PREVISIONI ALTERNATIVE

Guido Gambetta

Ottobre 1983

N. 3

Università di Bologna, Bologna

INDICE

Ringraziamo la dottessa Laura Parenti e il dott. Luciano Stefanini per la loro collaborazione alla stesura di una versione preliminare del presente lavoro.

1. Introduzione

Nelle applicazioni economiche è frequente il caso in cui si dispone, per una stessa variabile, di più previsioni alternative dei suoi valori futuri. Tali previsioni sono in generale ottenute con metodi diversi, quali l'uso di modelli econometrici, di modelli estrapolativi più o meno complessi, di giudizi di esperti. L'approccio tradizionale consiste nel cercare di individuare, per ogni applicazione, il metodo "migliore" in base a qualche criterio statistico. Un'alternativa interessante che è stata indagata recentemente, è quella di combinare in una previsione le previsioni alternative, calcolandone una media ponderata. Il problema è naturalmente quello di fissare un criterio per la determinazione dei pesi, tenendo conto che si combinano delle previsioni che, in linea di principio, sono basate su insiemi di informazione dipendenti fra di loro.

Consideriamo, per esempio, una serie storica X_t, $t = 1,\dots,T$ e supponiamo di essere interessati a prevedere il valore X_{t+1}. Supponiamo, altresì, di disporre di un numero $p > 1$ di previsioni di X_t per il periodo $t = n,\dots,T + 1$ ottenute con metodi alternativi.

Se indichiamo con $I_{t,p}$ l'insieme di informazioni di cui disponiamo, il problema sarà quello di calcolare il valore atteso condizionale

$$X^*_{T+1} = E\,(X_{T+1} \mid I_{t,p}) \qquad (1.1)$$

dove $\quad E(X_{T+1} - X^*_{T+1}) = E\,(e_{T+1}) = 0$

Nel nostro caso l'insieme di informazione sarà composta dalle p previsioni alternative

$$\hat{X}_{t,i} = E(X_t \mid I^i_t)$$

dove I^i_t è l'insieme di informazione sul quale è basata la i-ma previsione, con $i = 1,\dots, p$ e dai relativi errori di previsione

$$e_{t,i} = X_t - \hat{X}_{t,i}$$

dove $t = n,\dots, T$. Supponiamo anche che

$$E(e_{t,i}) = 0$$

$$E(e^2_{t,i}) = \sigma^2_i$$

Se si conoscessero i singoli insiemi di informazione I_t^i la soluzione ottimale al problema di trovare X^*_{T+1} sarebbe quella di combinarli in un unico insieme di informazione e di basare su questo la previsione cercata (vedi Reid (1974)). La situazione in cui è ottimale la (1.1) è evidentemente quella in cui le nostre conoscenze sono parziali e si riducono all'insieme I_p.

Tale situazione in pratica è molto frequente e comprende anche il caso in cui sarebbe troppo lungo e costoso costruire un modello per la previsione di X_{T+1} che incorpori gli insiemi I_t^i (costituiti in prevalenza dai singoli modelli di previsione).

La soluzione cercata sarà del tipo

$$X^*_{T+1} = \sum_{i=1}^{p} K_i \hat{X}_{T+1,i} \qquad (1.2)$$

con K_i pesi incogniti da determinare.

Nel prossimo paragrafo, considereremo la prima soluzione proposta da Bates e Granger (1969) e da Reid (1969). Nei paragrafi successivi suggeriremo una generalizzazione del primo approccio, introdotta soprattutto per ampliare il campo delle applicazioni pratiche della metodologia della combinazione di previsioni.

2. La determinazione dei pesi

La determinazione del peso da assegnare alle singole previsioni deve tenere conto almeno di due elementi: a) l'accuratezza previsiva dei singoli metodi; b) l'eventuale correlazione esistente fra gli errori di previsione degli stessi metodi.

Consideriamo la serie storica X_t e supponiamo, per semplicità, di avere per essa due previsioni alternative non distorte generate al tempo t per un periodo avanti. Siano $\hat{X}_{t,1}$, $\hat{X}_{t,2}$ tali previsioni con errori:

$$e_{t,i} = X_t - \hat{X}_{t,i} \qquad i=1,2$$

tali che:

$$E(e_{t,i}) = 0 \quad e \quad E(e^2_{t,i}) = \sigma^2_i \qquad i=1,2$$

e sia ρ il coefficiente di correlazione tra gli errori della prima serie di previsioni e quelli della seconda.

Una previsione combinata si può ottenere calcolando la media ponderata delle due previsioni individuali, assegnando ciò un peso K alla prima previsione ed un peso 1-K alla seconda; è stato proposto (vedi Bates e Granger (1969)) di determinare K in modo che la varianza dell'errore della combinazione

$$K_t^* = K\,\hat{X}_{t,1} + (1-K)\hat{X}_{t,2} \quad \text{sia minima}$$

L'errore della previsione X_t^* ottenuta dalla combinazione è:

$$e_t^* = X_t - X_t^* = Ke_{t,1} + (1-K)e_{t,2}$$

la sua varianza, che è funzione del peso K, avrà la seguente forma:

$$\sigma_*^2 = K^2\sigma_1^2 + (1-K)^2\sigma_2^2 + 2\rho K(1-K)\sigma_1\sigma_2 \qquad (2.1)$$

Si vuole scegliere K in modo che σ_*^2 sia minimo; quindi, derivando rispetto a K e annullando, si ottiene:

$$K_0 = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \qquad (2.2)$$

E' interessante notare che la (2.2) può essere riscritta nella seguente forma:

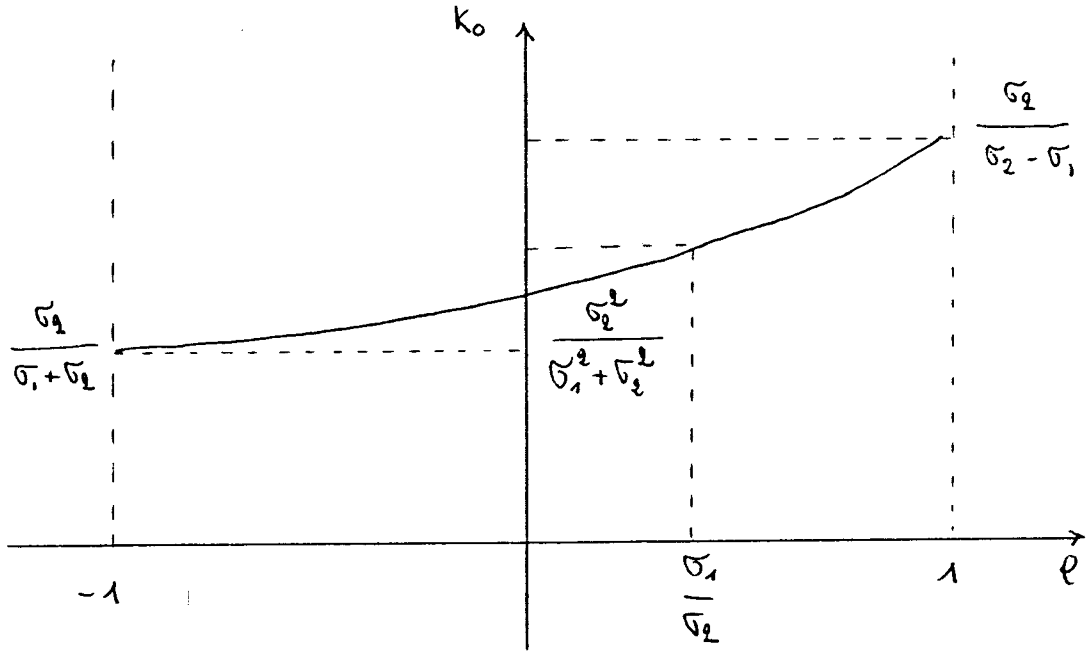
$$K_0 = \frac{\sigma_2(\sigma_2 - \rho\sigma_1)}{(\sigma_1 - \rho\sigma_2)^2 + \sigma_2^2(1-\rho^2)} \qquad (2.3)$$

Essendo il denominatore di (2.3) sempre positivo il segno di K_0 dipende solo dal numeratore.

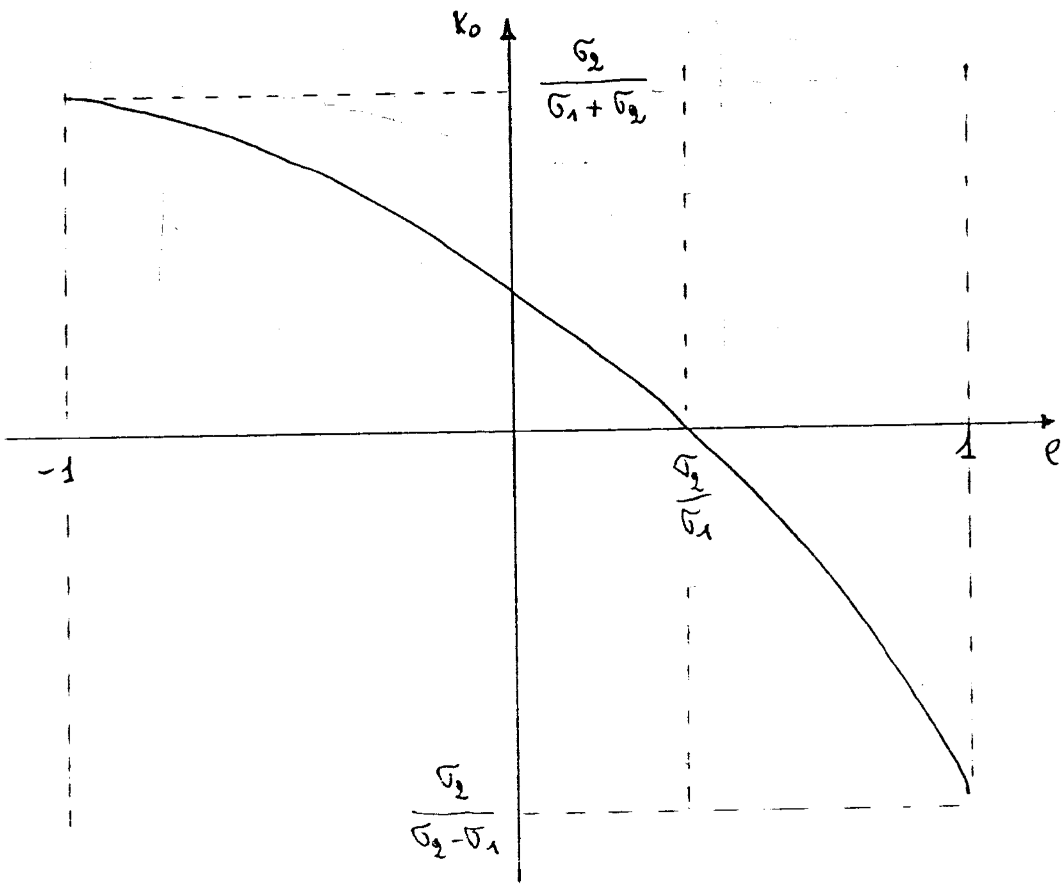
E' opportuno distinguere due casi:

a) Se $\sigma_1 < \sigma_2$, K_0 assume valori crescenti e sempre positivi, ma
   $1 - K_0 < 0$ per $\rho > \frac{\sigma_1}{\sigma_2}$

(vedi fig. 1)



b) Se invece $\sigma_2^2 < \sigma_1^2$ possiamo dire che K_0 è una funzione sempre decre- scente di ρ ed assume valori positivi per $-1 < \rho < \frac{\sigma_2}{\sigma_1}$ e valori negativi per $\frac{\sigma_2}{\sigma_1} < \rho < 1$ (vedi fig. 2):



In pratica, i valori dei pesi saranno spesso entrambi positivi. Tuttavia, nel caso di una correlazione molto alta fra gli errori di previsione, ci si può aspettare che le due previsioni sottostimino o sovrastimino entrambe il valore vero: allora viene dato un peso negativo alla previsione meno precisa nel tentativo di fare "indietreggiare" quella più precisa verso il valore vero.

Nel caso, infine, in cui si supponga $\sigma_1^2 = \sigma_2^2$, cioè che due previsioni abbiano la stessa "bontà previsiva", i pesi che vengono loro attribuiti nella combinazione sono entrambi uguali a $\frac{1}{2}$.

Sostituendo il valore ottimo di K in (2.1) otterremo il minimo valore della varianza dell'errore della combinazione:

$$\sigma_{*,0}^2 = \frac{\sigma_1^2 \sigma_2^2 (1 - \rho^2)}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

E' facilmente dimostrabile che $\sigma_{*,0}^2 \leq \min(\sigma_1^2, \sigma_2^2)$ cioè che la varianza degli errori della combinazione è in genere minore o uguale alla minima varianza degli errori delle singole previsioni. Infatti si ha:

$$\sigma_{*,0}^2 - \sigma_1^2 = \frac{-\sigma_1^2 (\sigma_1 - \rho\sigma_2)^2}{(\sigma_1 - \rho\sigma_2)^2 + \sigma_2^2(1 - \rho^2)}$$

che è chiaramente una quantità minore di 0 e perciò $\sigma_{*,0}^2 < \sigma_1^2$ e per simmetria $\sigma_{*,0}^2 < \sigma_2^2$. Vale l'uguaglianza cioè $\sigma_{*,0}^2 = \sigma_1^2$ per $\rho = \frac{\sigma_1}{\sigma_2}$ e $\sigma_{*,0}^2 = \sigma_2^2$ per $\rho = \frac{\sigma_2}{\sigma_1}$.

Può essere interessante anche studiare il comportamento di $\sigma_{*,0}^2$ in funzione della correlazione ρ soprattutto quando ρ si avvicina ai valori - 1 e 1.

Nel primo caso si ha che $\sigma_{*,0}^2$ tende a 0 e si ottiene così una combinazione perfetta con varianza nulla. Anche quando ρ tende a +1 la funzione $\sigma_{*,0}^2$ tende a 0, eccettuato il caso in cui $\sigma_1 = \sigma_2$ nel quale $\sigma_{*,0}^2$ ha per limite σ_1^2, infatti:

$$\lim_{\rho \to 1} \frac{\sigma_1^4 (1-\rho^2)}{2\sigma_1^2(1-\rho)} = \lim_{\rho \to 1} \frac{\sigma_1^2(1+\rho)}{2} = \sigma_1^2$$

Il fatto che la varianza degli errori della combinazione tenda a 0 è abbastanza intuitivo nel caso in cui ρ si avvicina a -1, lo è molto meno nel caso in cui ρ approssima 1; bisogna pensare, però, che ρ rappresenta la correlazione tra le serie degli errori delle due previsioni e non tra le previsioni stesse. Consideriamo, per esempio, il caso in cui $e_{t,2} = Ae_{t,1}$ dove A è una certa costante positiva. Si dimostra facilmente che in questo caso la serie dei valori osservati è una combinazione lineare esatta delle due previsioni considerate.

Essendo:

$$e_{t,1} = X_t - \hat{X}_{t,1} \qquad e_{t,2} = Ae_{t,1} = A(X_t - \hat{X}_{t,1})$$

la seconda previsione sarà della forma:

$$\hat{X}_{t,2} = X_t - e_{t,2} = X_t - A(X_t - \hat{X}_{t,1}) = (1-A)X_t + A\hat{X}_{t,1}$$

Allora supponendo $A \neq 1$ la serie da prevedere sarà:

$$X_t = -\frac{A}{1-A}\hat{X}_{t,1} + \frac{1}{1-A}\hat{X}_{t,2}$$

Gli stessi risultati possono essere ottenuti come caso particolare di un approccio Bayesiano più generale, come ha mostrato Winkler (1981). Se gli errori di previsione sono distribuiti in modo normale, con media zero e matrice di varianze-covarianze

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

e se abbiamo una distribuzione iniziale diffusa per X_t, allora la sua distribuzione finale sarà normale, con media X_t^* e varianza σ_*^2.

Il problema ha, così, una soluzione teorica soddisfacente. Nelle applicazioni pratiche, tuttavia, non si conoscono in generale i valori di σ_1^2, σ_2^2, e ρ, che devono pertanto essere stimati in qualche modo. Chiamando K_T il peso da attribuire al tempo T nella combinazione alla prima previsione e con $1-K_T$ quello da assegnare alla seconda, sono stati proposti alcuni metodi che forniscono K_T stimando σ_1^2, σ_2^2 e ρ in funzione degli errori passati delle due previsioni, denotati con

$\hat{e}_{1,1} \dots \hat{e}_{T-1,1}$; $\hat{e}_{1,2} \dots \hat{e}_{T-1,2}$

Solo K_1 può essere scelto arbitrariamente, solitamente si tende a prenderlo pari a 0,5 a meno che non si voglia privilegiare a priori un modello previsivo sull'altro.

Una naturale scelta per K_T è:

$$K_T = \frac{\sum_{t=T-n}^{T-1} (\hat{e}_{t,2}^2 - \hat{e}_{t,1}\hat{e}_{t,2})}{\sum_{t=T-n}^{T-1} (\hat{e}_{t,1}^2 + \hat{e}_{t,2}^2 - 2\hat{e}_{t,1}\hat{e}_{t,2})} \tag{2.4}$$

in cui si è posto:

$$\hat{\sigma}_1^2 = \sum_{t=T-n}^{T-1} \hat{e}_{t,1}^2 ; \quad \hat{\sigma}_2^2 = \sum_{t=T-n}^{T-1} \hat{e}_{t,2}^2 ; \quad \hat{\rho} = \frac{\sum_{t=T-n}^{T-1} \hat{e}_{t,1}\hat{e}_{t,2}}{\sqrt{\sum_{t=T-n}^{T-1} \hat{e}_{t,1}^2}\sqrt{\sum_{t=T-n}^{T-1} \hat{e}_{t,2}^2}}$$

La scelta del valore di n dipende dal numero di previsioni precedenti di cui si vuole tener conto, che può non essere uguale al numero massimo di quelle disponibili. Ci sono, certamente, dei casi in cui è sensato privilegiare le informazioni più recenti, perchè ci possono essere state delle variazioni strutturali che hanno reso gli errori di previsione passati poco significativi. In ogni caso la scelta di n è guidata da informazioni estranee ai metodi di previsione resi disponibili al ricercatore da altre fonti.

K_T definito da (2.4) può essere considerato anche come lo stimatore del peso K ottenuto con il metodo dei minimi quadrati basato sui dati disponibili dei periodi T-n,...,T-1.

La combinazione delle due previsioni al tempo t può essere riscritta nel seguente modo:

$$X_t^* - \hat{X}_{t,2} = K(\hat{X}_{t,1} - \hat{X}_{t,2})$$

ed equivalentemente:

$$e_{2,t} = K(e_{t,2} - e_{t,1}) + e_t^* \tag{2.5}$$

Perciò la stima di K nella (2.5) con il metodo dei minimi quadrati coincide esattamente con la (2.4).

L'approccio appena visto, basato sulla minimizzazione della varianza dell'errore della combinazione cercata, può essere esteso anche a più previsioni.

Supponiamo di avere p previsioni, un passo avanti, relative ad uno stesso valore X_T ottenute con p metodi alternativi. Supponiamo che siano non distorte. Si può considerare un vettore p-dimensionale $\hat{X}_T = (\hat{X}_{T,1}, \ldots, \hat{X}_{T,p})$ dove $\hat{X}_{T,i}$ è la i-esima previsione al tempo T.

La combinazione $X_T^* = K_T'\hat{X}_T$ sarà non distorta essendolo ogni singola previsione. Il problema è quindi di determinare $K_T' = (K_{T,1}, \ldots, K_{T,p})$ con il vincolo $K_T' \cdot 1 = 1$ dove $1' = (1, \ldots, 1)$ e $K_{T,i}$ rappresenta il peso da attribuire all'i-esima previsione nella combinazione. K_T sarà funzione delle varianze degli errori dei singoli modelli e delle loro correlazioni.

Si può dimostrare che la varianza dell'errore della combinazione è minimizzata da:

$$K_T = \frac{(\Sigma^{-1} \cdot 1)}{(1' \cdot \Sigma^{-1} \cdot 1)} \tag{2.6}$$

dove $\Sigma = E(e_T \, e'_T)$ con $e_T = X_T \cdot 1 - P_T$.

La (2.6) per p = 2 può essere facilmente ricondotta alla (2.2).

Poichè, in pratica, non si conosce la matrice Σ di varianza e covarianza degli errori, calcoleremo

$$(\hat{\Sigma})_{ij} = n^{-1} \sum_{t=T-n}^{T-1} \hat{e}_{t,i} \, \hat{e}_{t,j} \text{ si ha}$$

$$\hat{K}_T = \frac{(\hat{\Sigma}^{-1} \cdot 1)}{(1' \cdot \hat{\Sigma}^{-1} \cdot 1)}$$ che corrisponde al vettore degli stimatori

dei parametri della regressione della serie storica sulle previsioni ottenute con i diversi metodi.

L'approccio fin qui considerato ottiene i pesi ottimali della combinazione delle previsioni presupponendo due ipotesi fondamentali oltre a quelle già elencate:

1) che il costo associato all'errore di previsione sia calcolato con una funzione quadratica del tipo

$$C(e_t) = \alpha e_t^2 \qquad \alpha > 0 \tag{2.7}$$

2) che la varianza $\sigma_i^2 = E(e_{t,i}^2)$ dei singoli errori di previsione sia costante nel tempo.

Recentemente, Engle, Granger e Kraft (1982), hanno proposto un metodo per tener conto del fatto che la matrice di covarianza degli errori di previsione può variare nel tempo.
Può essere anche interessante mettere in discussione la scelta della funzione di costo quadratica, che risulta inappropriata in alcune applicazioni pratiche.

Vi sono almeno due strade lungo le quali generalizzare la funzione di perdita: a)cercare funzioni asimmetriche che attribuiscono un peso diverso a errori uguali in valore assoluto, ma di segno opposto; b) mantenere la simmetria della funzione, ma abbandonare l'ipotesi di forma strettamente quadratica.

Nel prossimo paragrafo tratteremo alcuni esempi di entrambe le situazioni e ne illustreremo le implicazioni pratiche per il problema che stiamo considerando.

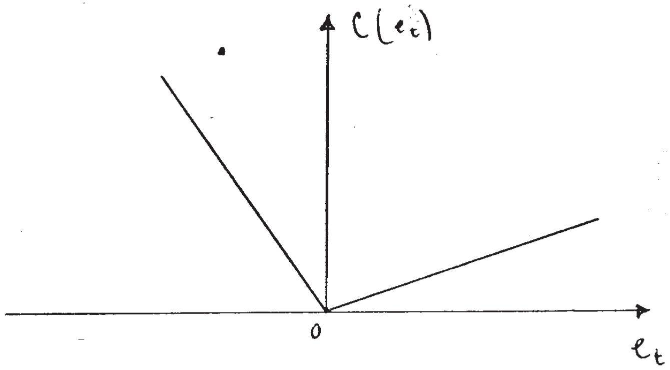
3. La scelta della funzione di perdita.

Nell'ambito della ricerca applicata la previsione di una variabile quantitativa non è quasi mai un obiettivo finale, ma soltanto un passo di un processo decisionale complesso.

In altre parole, il valore previsto della variabile entra come uno degli elementi di informazione su cui si basa una decisione. A volte è possibile valutare i costi delle conseguenze di tali decisioni e una parte di tali costi sarà direttamente proporzionale all'entità degli errori di previsione commessi. Si possono certamente immaginare dei casi in cui la funzione di costo appropriata è diversa dalla (2.7), per esempio quando è necessario introdurre delle asimmetrie.

Consideriamo il caso in cui una sottostima delle vendite da parte di una impresa comporta una perdita di clienti e una sovrastima, invece, una accumulazione indesiderata di scorte. Non vi è nessuna ragione per ritenere che i costi siano simmetrici in tale situazione.

Un caso interessante di funzioni non simmetriche è quello di funzioni "spline", che sono funzioni polinomiali a tratti con punti di discontinuità chiamati "nodi". L'uso di tali funzioni in problemi di econometria e, in particolare, nella specificazione di funzioni di perdita è illustrato da Poirier (1976).
Un esempio di tali funzioni nel caso lineare è



dove il punto di discontinuità è per $e_t = 0$.

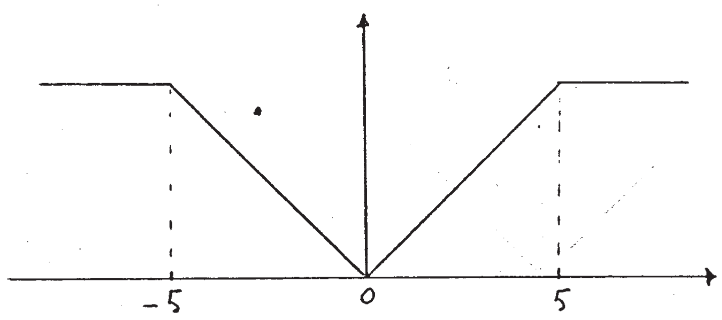
Un altro caso che vogliamo considerare è quello di funzioni che possono eventualmente essere simmetriche come caso particolare, ma in cui vogliamo trattare in modo diverso dagli altri errori di previsione che, per qualche motivo, riteniamo anomali e irripetibili.
Consideriamo, per esempio, il comportamento di due previsori:

t	$X_t-\hat{X}_{t,1}$	$X_t-\hat{X}_{t,2}$
1	0	4
2	0	6
3	-12	-5
4	0	3
5	0	4

Il metodo dei minimi quadrati privilegerebbe il secondo previsore per il quale $\sum_{t=1}^{5} e^2_{t,2}=102$, rispetto al primo che ha $\sum_{t=1}^{5} e^2_{t,1}=144$.

Ora, se abbiamo qualche ragione per credere che nel periodo t=3 sia accaduto qualche fatto particolare (scioperi, fattori stagionali, ecc.) per il quale possa essere trattato separatamente, potremmo essere interessati a dare un peso maggiore alla previsione che non commette errori nei periodi "normali". Ciò può essere ottenuto in modo generale con procedure "robuste" (vedi in particolare Huber (1981)) che utilizzano funzioni di perdita non quadratiche.

Una funzione appropriata nel caso considerato più sopra potrebbe essere



che attribuirebbe lo stesso peso a tutti gli errori maggiori di 5 in valore assoluto. (Per un altro caso di utilizzazione di una tale funzione di costo vedi Granger (1969)).

Introduciamo, allora, due funzioni di perdita generali che hanno, rispettivamente, come caso particolare e come caso limite, la usuale funzione quadratica (2.7).

La prima funzione considerata è una funzione spline quadratica del tipo:

$$C_A(e_t) = \begin{cases} \alpha_1 e_t^2 & \text{se } e_t \geq 0 \\ \alpha_2 e_t^2 & \text{se } e_t < 0 \end{cases} \qquad \alpha_1, \alpha_2 > 0 \qquad (3.1)$$

della quale la (2.7) è un caso particolare per $\alpha_1 = \alpha_2 = \alpha$.

La seconda funzione viene introdotta per trattare, oltre ai problemi di asimmetria, anche quelli dell'esistenza di valori "anomali". Una funzione interessante si è dimostrata la seguente:

$$C_E(e_t) = \begin{cases} \dfrac{E^2}{2\alpha_1^2} \quad 1 - e^{-\left(\frac{\alpha_1 e_t}{E}\right)^2} & \text{se } e_t \geq 0 \\[2ex] \dfrac{E^2}{2\alpha_2^2} \quad 1 - e^{-\left(\frac{\alpha_2 e_t}{E}\right)^2} & \text{se } e_t < 0 \end{cases} \qquad (3.2)$$

dove $\alpha_1, \alpha_2 > 0$

Quando $\alpha_1 = \alpha_2 = \alpha$ e il valore del parametro E è abbastanza grande i risultati ottenuti con tale funzione coincidono praticamente con quelli ottenuti con la (2.4).

I grafici delle due funzioni nel caso più generale sono riportati, rispettivamente, nelle figure 3 e 4.

Il calcolo della combinazione ottimale di previsioni richiede che sia nota la funzione di perdita, in particolare, che si conoscano i valori di α_1, α_2 ed E. Nelle applicazioni pratiche dovrebbe essere possibile calcolare, almeno in modo approssimato, il costo associato a errori di previsione di differente entità, disponendo così di alcune osservazioni della funzione di perdita. Con una semplice interpolazione delle funzioni $C_A(e_t)$ o $C_E(e_t)$ fra tali punti si potrà, allora, determinare i valori di α_1, α_2 ed E.
Nei casi in cui ciò non fosse possibile, si potrà procedere con considerazioni di tipo "qualitativo" dettate dalla natura del problema o con un'analisi di sensitività effettuando prove alternative con

differenti valori dei parametri. In base ad una logica di massima verosimiglianza, si potrebbe, in quest'ultimo caso, scegliere quella combinazione di previsioni che risulta migliore in base a qualche criterio statistico (per esempio lo scarto quadratico medio fra valori osservati e valori previsti sul periodo campione).

A conclusione ricordiamo un suggerimento che può essere utile per l'uso della seconda funzione: dato che $C_E(e_t)$ ha i punti di flesso per $e_t = \pm \frac{E}{\alpha\sqrt{2}}$ soltanto errori di previsione maggiori di tale valore inizieranno ad essere trattati come anomali. Sarà pertanto opportuno controllare che i valori di E e di α (oppure α_1 e α_2) siano tali da definire correttamente tali errori anomali.

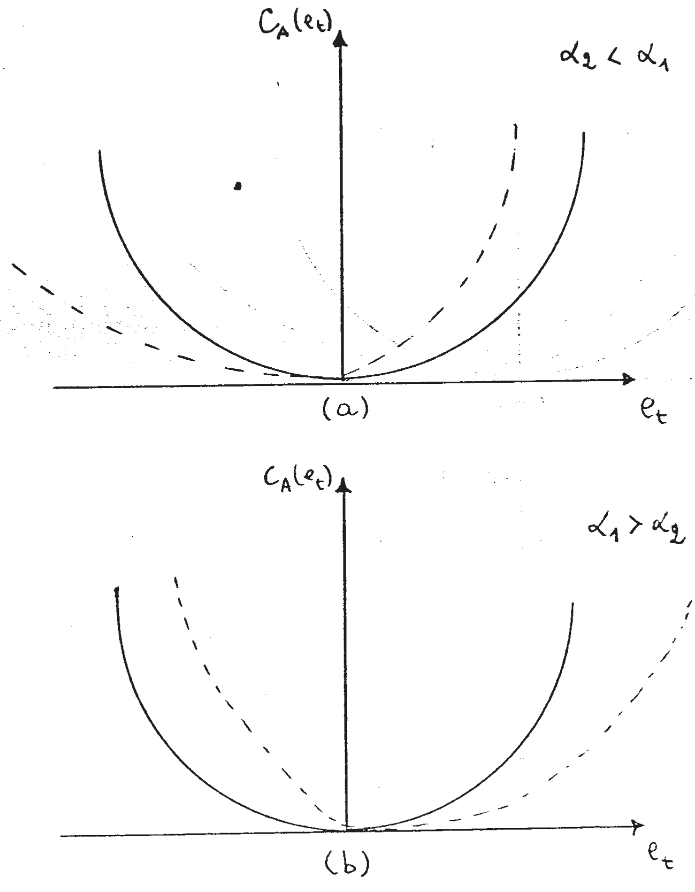


fig. 3

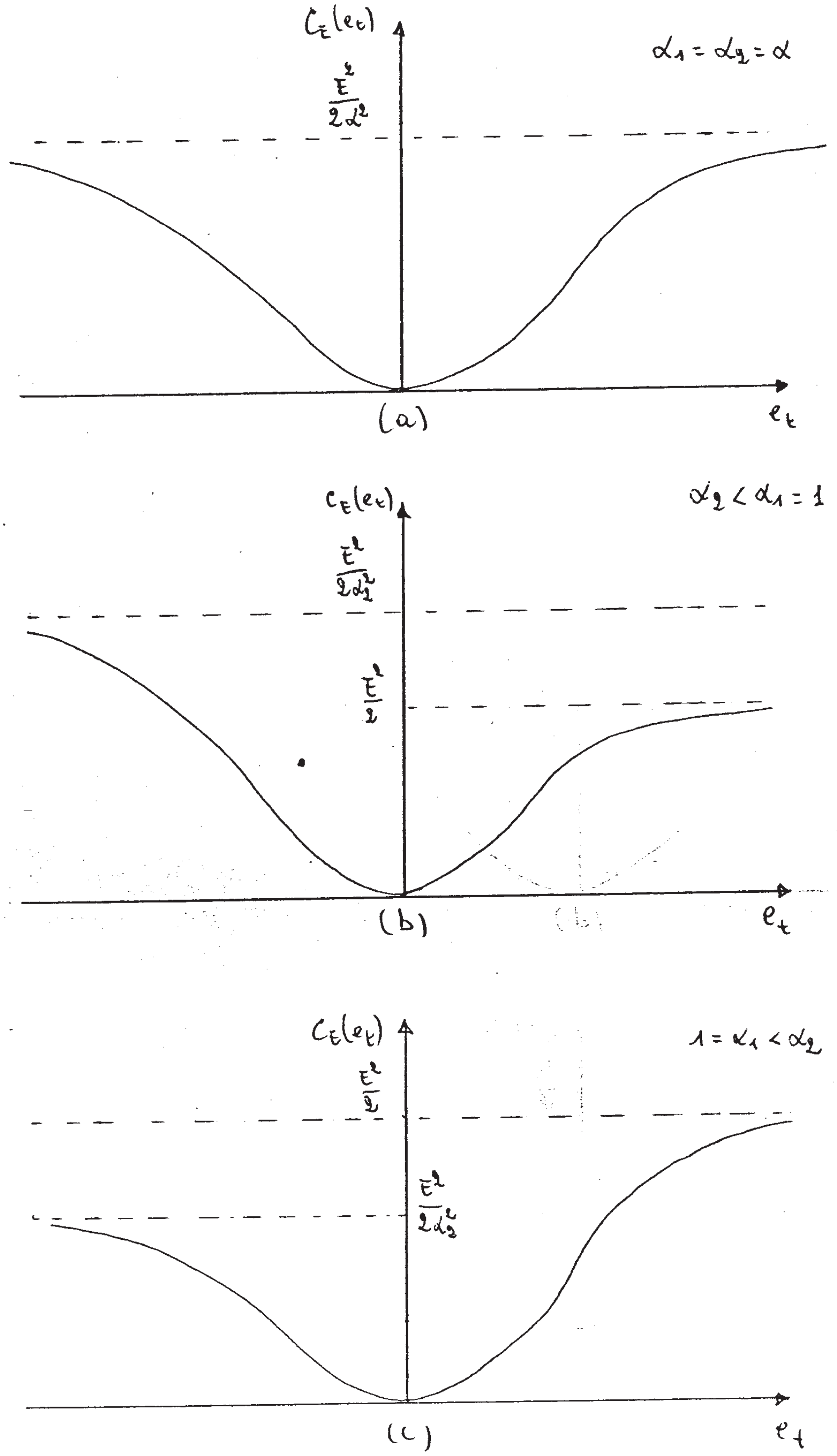

$C_E(e_t)$
$\alpha_1 = \alpha_2 = \alpha$
$\frac{E^2}{2\alpha^2}$
(a)
e_t
$C_E(e_t)$
$\alpha_2 < \alpha_1 = 1$
$\frac{E^2}{2\alpha_2^2}$
$\frac{E^2}{2}$
(b)
e_t
$C_E(e_t)$
$1 = \alpha_1 < \alpha_2$
$\frac{E^2}{2}$
$\frac{E^2}{2\alpha_2^2}$
(c)
e_t

fig. 4

4. Osservazioni conclusive.

Abbiamo considerato alcuni metodi per determinare i pesi di una combinazione "ottima" di previsioni. Tale combinazione non è necessariamente convessa, anche se i pesi sommano a uno, in quanto alcuni possono essere negativi. Potrebbe, invece, essere interessante impostare la scelta del previsore ottimale in termini decisionali. Ad ogni vettore di pesi $K = (K_1, K_2, \ldots, K_p)$ con componenti non negative e di somma unitaria si può attribuire allora il significato di distribuzione di probabilità sui p casi possibili riguardanti il sorteggio di una tra le p assegnate decisioni, ognuna delle quali corrisponde alla scelta di un particolare previsore $\hat{X}_{t,i}$. Una combinazione lineare dei previsori corrisponde a una decisione mista; una decisione pura è una particolare decisione mista a cui è associato un vettore K con una sola componente positiva e uguale a 1.

La scelta di una combinazione convessa può essere attuata anche nell'approccio Bayesiano ricordato nel secondo paragrafo. In tal caso, però, la distribuzione finale non sarà più una distribuzione normale, ma una mistura delle singole distribuzioni normali sugli errori di previsione.

Certamente l'argomento merita di essere sviluppato e soprattutto arricchito con applicazioni pratiche, sulla scia dei risultati incoraggianti ottenuti da alcuni autori. Utilizzando diversi metodi per determinare i pesi. Newbold e Granger (1974), Makridakis e al. (1982) e Makridakis e Winkler (1983) trovano che una combinazione di previsioni risulta più accurata delle singole previsioni e tende a ridurre la variabilità degli errori di previsione. Alcuni esperimenti preliminari condotti su serie economiche italiane hanno confermato tali risultati.

Bibliografia

Bates, J.M., Granger, C.W.J., (1969), "The combination of forecasts", Operational Research Quarterly, 4, pp. 451-468.

De Groot, M.H. (1974), "Reaching a consensus", Journal of the American Statistical Association, 69, pp. 118-121.

Engle, R.F., Granger, C.W.J., Kraft, D. (1982), "Combining competing forecasts of inflation using a bivariate ARCH model", University of California, San Diego, (dattiloscritto).

Granger, C.W.J. (1969), "Prediction with a generalized cost of error function", Operational Research Quarterly,20, pp. 199-207.

Huber, P.J. (1981), "Robust Statistics", New York, J. Wiley e Sons.

Makridakis, S. e al. (1982), "The accuracy of extrapolation (Time Series) methods: results of a forecasting competition", Journal of Forecasting, 1, pp. 111-153.

Makridakis, S., Winkler, R.L. (1983), "The combination of forecasts", Journal of the Royal Statistical Society, A, 146, pp.150-157.

Morris, P.A. (1977),"Combining expert judgments: a Bayesian approach", Management Science, 23, pp. 679-693.

Newbold P., Granger, C.W.J. (1974), "Experience with forecasting univariate time series and the combination of forecasts", Journal of the Royal Statistical Society, A, 137, pp. 131-164.

Poirier, D.J. (1976), The econometrics of structural change, Amsterdam, North Holland.

Reid, D.J. (1974), "Discussion of the paper by P. Newbold e C.W.J. Granger, "Experience with forecasting univariate time series and the combination of forecasts", Journal of the Royal Statistical Society, A, 137, pp. 146-148.

Reid, D.J. (1969), "A comparative study of time series prediction techniques on economic data", Ph.D. Thesis, University of Nottingham.

Winkler, R.L. (1968), "The consensus of subjective probability distributions", Management Science, 15, pp. B61-B75.

Winkler , R.L. (1981), "Combining probability distributions from dependent information sources", Management Science, 27, pp. 479-488.