

**Il programma “STIMA RAPIDA”:
nota metodologica ed alcune applicazioni
alle indagini congiunturali sulle imprese**

Roberto Gismondi (*)

(*) *ISTAT - Servizio SCO*

Sommario

In numerosi campi della ricerca applicata si presenta spesso il problema di dover predisporre stime “rapide” relativamente a fenomeni quantitativi il cui ammontare definitivo sarà noto con certezza solo dopo un certo periodo, successivo a quello entro il quale tali stime si rendono necessarie. Tra i tanti criteri utilizzabili a tale fine, quelli proposti nel programma “Stima_rapida.xls” derivano da 2 particolari stimatori, ottimali sotto determinate condizioni, oltre al più tradizionale metodo del quoziente. Il programma può essere utilizzato per stimare un ammontare al tempo t , sia supponendo noto l’ammontare al tempo $(t-k)$; sia sfruttando i dati (disponibili) relativi al tempo t e *ad una parte dell’intera popolazione* di riferimento (un *panel*, un sub-campione naturale di rispondenti rapidi, ecc.). Una peculiarità del programma consiste nel fatto che possono essere ottenuti risultati soddisfacenti senza la necessità di disporre di serie storiche particolarmente lunghe. In teoria potrebbero essere sufficienti 2 o 3 coppie di osservazioni. Riguardo all’efficienza delle procedure proposte, ovviamente non possono essere garantiti risultati qualitativamente comparabili con quelli derivati da metodi più sofisticati basati sull’analisi delle serie storiche, che restano certamente preferibili in presenza di serie storiche lunghe. Ciò nonostante i risultati delle numerose simulazioni proposte nel paragrafo 6 evidenziano la possibilità di utilizzare una “batteria” di metodi, in funzione del ritardo tra l’ultimo periodo osservato disponibile ed il periodo oggetto di stima e delle caratteristiche intrinseche della variabile da stimare.

1. Necessità di stime “rapide” in presenza di serie storiche brevi¹

In numerosi campi della ricerca applicata si presenta spesso il problema di dover predisporre stime “rapide” relativamente a fenomeni quantitativi il cui ammontare definitivo sarà noto con certezza solo dopo un certo periodo, successivo a quello entro il quale tali stime si rendono necessarie.

Tra i tanti criteri utilizzabili a tale fine, quelli proposti nel programma “Stima_rapida.xls” derivano dagli stimatori generali riportati nelle formule [1] e [6] dei paragrafi seguenti. Essi si basano sull’ipotesi di poter disporre di una anticipazione relativamente al fenomeno di interesse, derivabile a sua volta da queste due possibili situazioni:

- per stimare l’ammontare al tempo t , si sfrutta la conoscenza dell’ammontare noto al tempo $(t-k)$: un esempio è dato dalla necessità di stimare l’ammontare che una data variabile assumerà alla fine dell’anno, essendo noto l’ammontare dei primi n mesi del medesimo anno, con $n < 12$;
- per stimare l’ammontare al tempo t , si sfruttano i dati (disponibili) relativi al tempo t e *ad una parte dell’intera popolazione* di riferimento (un *panel*, un sub-campione naturale di rispondenti rapidi, un sottoinsieme di unità i cui dati sono disponibili anticipatamente): un esempio è dato, come sopra, dalla necessità di stimare l’ammontare che una data variabile assumerà alla fine dell’anno per l’intera popolazione di riferimento (ad esempio, intera Italia), essendo noto l’ammontare complessivo che tale variabile ha assunto per l’intero anno ma con riferimento ad una parte della popolazione di riferimento (ad esempio, una certa regione).

Sebbene il programma sia utilizzabile in entrambe le situazioni, la sua utilità maggiore si riferisce, con tutta probabilità, al secondo dei casi suddetti, che consente di sfruttare tutte le informazioni “preliminari” (o “anticipatorie”) di cui si dispone al momento di dover calcolare la stima.

La sua peculiarità consiste nel fatto che possono essere ottenuti risultati soddisfacenti senza la necessità di disporre di serie storiche particolarmente lunghe. In teoria potrebbero essere sufficienti 2 o 3 coppie di osservazioni.

Riguardo all’efficienza delle procedure proposte, ovviamente non possono essere garantiti risultati qualitativamente comparabili con quelli derivati da metodi più sofisticati basati sull’analisi

¹ “Prevedere il futuro è abbastanza impossibile, ma non tentare di prevederlo è dannoso perché le previsioni, anche se sbagliate, servono a far ragionare coloro che, dopo qualche tempo e a giusta ragione, prenderanno in giro i profeti” (Bertrand Russell).

delle serie storiche, che restano certamente preferibili in presenza di serie storiche lunghe. In particolare, non è esplicitamente prevista la possibilità di sfruttare eventuali variabili esplicative, sebbene i criteri della stima ottimale del livello e della variazione descritti nei due paragrafi seguenti possano essere impostati in modo da sfruttare una singola variabile esplicativa².

Ciò nonostante i risultati delle numerose simulazioni proposte nel paragrafo 6 evidenziano la possibilità di utilizzare una “batteria” di metodi, in funzione del ritardo tra l’ultimo periodo osservato disponibile ed il periodo oggetto di stima e delle caratteristiche intrinseche della variabile da stimare.

Al riguardo si premette che il programma prevede il calcolo di 18 stimatori, derivati dall’applicazione di tre metodologie di stima che presentano diverse opzioni operative, legate soprattutto alla tecnica di stima dei parametri ed all’utilizzo o meno degli errori di stima precedenti, sebbene su tali aspetti si rimandi al paragrafo 5. Nei paragrafi 2 e 3 ci si occuperà separatamente dei problemi della stima del livello e della variazione, come suggerito anche in Rao, Srinath e Quenneville (1989) ed in lavori precedenti (Gismondi, 1996 e 1998), mentre nel paragrafo 4 si propongono alcuni criteri per la stima dei parametri incogniti necessari per l’implementazione degli stimatori.

2. Stima del livello

Si supponga, con riferimento ad una variabile quantitativa non negativa N , di considerare due periodi consecutivi, indicati con i pedici 1 e 2, e che tramite una misurazione preliminare indicabile con il pedice P - basata su un sottoinsieme del totale delle unità statistiche da osservare o su una misurazione riferita ad un periodo precedente - siano disponibili gli ammontari preliminari N_{1P} e N_{2P} , riferiti rispettivamente ai tempi 1 e 2; solo in un tempo successivo si potranno conoscere gli esatti ammontari finali della variabile oggetto di studio³ in entrambi i tempi, ossia N_1 e N_2 .

² Nel successivo stimatore definito dalla [1] si potrebbe utilizzare tale variabile esplicativa come variabile preliminare, sebbene possa risultare necessario effettuare una opportuna traslazione preventiva dell’unità di misura con cui è espressa per renderla sempre non superiore alla variabile finale oggetto di stima.

³ Ad esempio, se si fosse interessati alla stima del numero di imprese attive ai tempi 1 e 2 relativamente alla divisione 52 della classificazione ATECO (commercio al dettaglio), si potrebbe supporre di conoscere l’ammontare di imprese attive in un particolare gruppo (ad esempio, il gruppo 52.2 delle imprese specializzate con vendita prevalente di alimentari e bevande), oppure delle imprese attive in una certa regione, o ancora di un *panel* la cui mortalità interna - in termini di imprese che rimangono attive nel tempo - possa comunque essere utile per stimare l’ammontare complessivo delle imprese in essere. In ogni caso, il sottoinsieme di osservazioni si supporrà disponibile in tempi assai ridotti rispetto a quelli necessari per poter stabilire l’effettiva consistenza di imprese attive nell’intero universo, per cui con riferimento ad esso si parlerà di osservazioni *preliminari*.

Supponendo di trovarci al tempo 2, e di disporre sia delle stime preliminari, sia degli ammontari finali relativi al tempo 1, nonché della stima preliminare relativa allo stesso tempo 2, il problema consiste nel determinare uno stimatore dell'ammontare finale incognito relativo al tempo 2 che risulti "ottimale", sulla base delle informazioni disponibili, nonché delle ipotesi e delle specificazioni formali che saranno illustrate nel prosieguo. Lo stimatore del livello corrente N_2 incognito sarà scrivibile in questa forma:

$$N_L = N_{2P} + \alpha(N_1 - N_{1P}) \quad [1]$$

dependente dal parametro incognito α . In particolare, se $\alpha = 0$ si ha $N_L = N_{2P}$, ossia uno stimatore che utilizza solo le informazioni derivate dalle osservazioni preliminari, mentre se $\alpha = 1$ si ha $N_L = N_1 + (N_{2P} - N_{1P})$, e le eventuali distorsioni dei due stimatori dipendono dai momenti primi delle grandezze coinvolte. Il problema consiste nella determinazione di un valore *ottimale* per α che minimizzi la media quadratica dell'errore (*MQE*) sotto condizioni non particolarmente restrittive. In questo contesto gli operatori di media e varianza saranno valutati in funzione della variabilità delle osservazioni misurabili sulle medesime unità del campione su cui è valutato l'ammontare finale Y , supponendo di poter replicare tali osservazioni più volte nello stesso istante di riferimento: secondo lo stesso approccio utilizzato da Rao, Srinath e Quenneville (1989), la variabilità non riguarda dunque un disegno campionario, ma il modello di *superpopolazione* sottostante alle modalità di risposta delle singole unità. Si supporrà poi valida, d'ora in avanti, l'ipotesi seguente:

ipotesi 1:

$$E(N_2 - N_{2P}) = \gamma E(N_1 - N_{1P}) = \gamma \delta .$$

In altri termini, si suppone che al tempo 2 la sottostima (relativa al modello di risposta) dell'ammontare preliminare rispetto all'ammontare totale sia in media proporzionale all'omologa sottostima relativa al tempo 1. Cambiando la coppia di periodi successivi considerata (t e $t+1$) cambierà, in generale, anche il valore del parametro, per cui in effetti $\gamma = \gamma_t$ e la costanza di γ va dunque riferita, in assenza di ulteriori ipotesi, solo alla suddetta coppia di periodi in esame⁴. *In primis*, γ verrà supposto libero da vincoli, anche se si vedrà in seguito come risulti auspicabile (e

⁴ Ad esempio, se sulla precisione delle stime preliminari infraannuali influisse la stagionalità del fenomeno osservato, l'implementazione della stessa tecnica previsionale potrebbe condurre, a parità delle restanti condizioni, ad una distorsione in valore assoluto più elevata in alcuni periodi dell'anno per il solo effetto del diverso ordine di grandezza del fenomeno studiato. Il verificarsi di una diminuzione nel tempo del valore (assoluto) assunto dal parametro γ potrebbe interpretarsi come un progressivo miglioramento della precisione delle stime preliminari.

plausibile) l'ipotesi $0 \leq \gamma \leq 1$. Posto poi che le distorsioni dei valori finali N_1 e N_2 siano nulle⁵, ossia che $B(N_1) = B(N_2) = 0$, segue dalla [1] che:

$$E(N_L) = E(N_{2P}) + \alpha E(N_1 - N_{1P}) = E(N_{2P}) + \alpha \delta$$

cioè:

$$|B(N_L)| = |E(N_{2P}) - E(N_2) + \alpha \delta| = |-\gamma \delta + \alpha \delta| = |\alpha - \gamma| \delta \quad [2]$$

e quindi la posizione $\alpha = \gamma$ minimizzerebbe, rendendola nulla, la distorsione dello stimatore del livello [1].

Privilegiando la scelta del valore ottimale di α nell'ottica della minimizzazione dell'errore quadratico medio, occorre ricavarne l'espressione formale e procedere alla sua minimizzazione con il metodo dei moltiplicatori di Lagrange⁶. Calcolando l'espressione formale della varianza della [1] e ricordando che, come noto, $MQE(N_L) = V(N_L) + [B(N_L)]^2$, dove V è l'operatore varianza, si ottiene dopo alcuni passaggi:

$$\begin{aligned} MQE(N_L) &= V(N_{2P}) + \alpha^2 V(N_1 - N_{1P}) + 2\alpha COV[N_{2P}(N_1 - N_{1P})] = \\ &= V(N_{2P}) + \alpha^2 V(N_1) - \alpha^2 V(N_{1P}) + 2\alpha^2 COV(N_1; N_{1P}) + \\ &\quad \underline{+ 2\alpha COV(N_1; N_{2P}) - 2\alpha COV(N_{1P}; N_{2P})} + (\alpha - \gamma)^2 \delta^2 . \end{aligned}$$

Una semplificazione della formulazione precedente è garantita dalla seguente:

ipotesi 2: se $N_1 = N_{1P} + Z_1$ \rightarrow $COV(N_{1P}; Z_1) = COV(N_{2P}; Z_1) = 0$.

In altri termini si suppone di poter scomporre ogni ammontare incognito nella somma di due componenti, date dall'osservazione preliminare N_P e dalla componente non osservabile preliminarmente Z , mutuamente incorrelate indipendentemente dal periodo di riferimento (1 o 2). Conseguentemente si avrà che:

⁵ Il simbolo B indica la distorsione (BIAS).

⁶ In ogni caso, si ribadisce come una stima di α basata solo sul confronto tra due periodi successivi invece che su un arco temporale più lungo comporta, in genere, una perdita di efficienza. D'altra parte, circa la frequente indisponibilità di serie storiche sufficientemente lunghe e comparabili si è già fatta menzione nella premessa: i motivi sono in genere strettamente inerenti al particolare campo di interesse (finalità stesse della rilevazione, cambiamenti nella tecnica di raccolta dei dati, variazione dei domini di studio, aspetti definatori, eventuali slittamenti di base, cambiamenti nella tecnica di correzione dei dati anomali, ecc.). Comunque, occorre sempre valutare preliminarmente il *trade/off* esistente tra maggiore semplicità di implementazione di uno stimatore e la relativa perdita di precisione.

$$COV(N_{1P}; N_1) = V(N_{1P}) \quad \text{e} \quad COV(N_{2P}; N_1) = COV(N_{2P}; N_{1P}).$$

E' allora evidente come le due covarianze sottolineate nella precedente espressione dell'errore quadratico medio dello stimatore del livello N_L si annulleranno e si avrà più semplicemente:

$$MQE(N_L) = V(N_{2P}) + \alpha^2 V(N_1) - \alpha^2 V(N_{1P}) + (\alpha - \gamma)^2 \delta^2$$

da cui derivano facilmente gli errori quadratici medi corrispondenti ai casi particolari $\alpha=1$ e $\alpha=0$. La derivata prima della precedente espressione è pari a:

$$2\alpha V(N_1) - 2\alpha V(N_{1P}) + 2(\alpha - \gamma) \delta^2$$

che uguagliata a zero fornisce questa soluzione:

$$\alpha_L^* = \frac{\gamma \delta^2}{V(N_1) - V(N_{1P}) + \delta^2} \quad [3]$$

Infine, non è difficile ricavare la conseguente formula per l'errore quadratico medio dello stimatore del livello:

$$MQE(N_L^*) = V(N_{2P}) + \gamma \delta^2 (\gamma - \alpha_L^*) \quad [4]$$

Si noti come, in pratica, risulti frequente il ricorso ad uno stimatore di più immediata implementazione basato sul metodo del quoziente⁷, dato da:

$$N_L = \frac{N_{2P}}{N_{1P}} N_1 \quad [5]$$

⁷ Il ricorso a questa tecnica presuppone, in genere, una forte rappresentatività delle osservazioni preliminari rispetto al totale incognito, o comunque un legame longitudinale molto stabile tra le due componenti, il che è presumibilmente verificabile solo nel breve periodo. D'altra parte, l'uso del valore ottimale del parametro α non comporta necessariamente una distorsione minore rispetto a quella dello stimatore quoziente, peraltro utilizzato dall'INSEE (l'istituto di statistica francese) per la stima degli indici provvisori delle vendite del commercio al minuto.

ottenibile ponendo nella [1] $\alpha_L = N_{2P}/N_{1P}$; una versione lievemente più sofisticata è ottenibile correggendo il precedente stimatore per tenere conto dell'errore di stima occorso tra i tempi 1 e 0, ossia calcolando:

$$N_L = \frac{N_{2P}}{N_{1P}} N_1 \frac{\left(\frac{N_1}{N_0}\right)}{\left(\frac{N_{1P}}{N_{0P}}\right)} \quad \text{ottenibile ponendo} \quad \alpha_L = \frac{N_{2P} \left[\left(\frac{N_1}{N_{1P}}\right)^2 \frac{N_{0P}}{N_0} - 1 \right]}{(N_1 - N_{1P})}.$$

3. Stima della variazione

Sulla base delle formalizzazioni e delle varie ipotesi introdotte nel paragrafo precedente si può così definire lo stimatore della variazione tra i due tempi, ossia della quantità incognita ($N_2 - N_1$):

$$N_V = N_L - N_1 \quad [6]$$

$$E(N_V) = E(N_L) - E(N_1) = E(N_{2P}) + \alpha\delta - E(N_1)$$

e quindi la distorsione di tale stimatore sarà data da:

$$|B(N_V)| = |E(N_V) - E(N_2) + E(N_1)| = |\alpha - \gamma|\delta. \quad [7]$$

Quindi lo stimatore della variazione tra i due tempi avrà la stessa distorsione dello stimatore del livello ed anche in questo caso la distorsione sarà nulla qualora $\alpha=\gamma$. Il calcolo dell'espressione formale della varianza dello stimatore [6], ottenibile sfruttando l'identità:

$$VAR(N_V) = VAR(N_L) + VAR(N_1) - 2COV(N_L, N_1)$$

è più complesso rispetto a quello relativo alla stima del livello, ma basandosi su ipotesi del tutto analoghe a quelle già introdotte viene per semplicità omissa. Dopo aver aggiunto a tale espressione della varianza il quadrato della distorsione e dopo aver calcolato ed uguagliato a zero la corrispondente derivata prima si ottiene, infine, il parametro α che rende minimo l'errore quadratico medio dello stimatore della variazione, dato da:

$$\alpha_V^* = 1 - \frac{(1-\gamma)\delta^2}{V(N_I) - V(N_{1P}) + \delta^2} = \alpha_L^* + \frac{V(N_I) - V(N_{1P})}{V(N_I) - V(N_{1P}) + \delta^2} \quad [8]$$

che quindi *non coincide* con il corrispondente peso ottimale per la stima del livello. La formula dell'errore sarà infine data da:

$$MQE(N_V^*) = [V(N_{2P}) + V(N_I) - 2COV(N_{1P}, N_{2P}) + \gamma^2 \delta^2] - [V(N_I) - V(N_{1P}) + \gamma \delta^2] \alpha_V^*. \quad [9]$$

Dalle espressioni formali degli errori quadratici medi dei due stimatori ottimali si deduce che, supposto $\gamma \cong 1$, se il quadrato della distorsione δ risultasse molto più elevato della differenza tra le varianze degli ammontari finale e preliminare del tempo 1, allora probabilmente $\alpha \cong 0$ costituirebbe la scelta ottimale, in quanto oltre a rendere nulla la distorsione minimizzerebbe anche l'errore quadratico medio, sia per la stima del livello che della variazione. D'altra parte, sia la crescita della distorsione δ che della varianza dell'ammontare incognito al tempo 1 relativo ai non rispondenti aumentano l'imprecisione di entrambe le stime.

Sulla base dei precedenti risultati va notato come la validità dell'ipotesi $0 \leq \alpha \leq 1$, sia per la stima del livello che della variazione equivale alle seguenti disuguaglianze algebriche:

$$0 \leq \alpha_L^* \leq 1 \leftrightarrow 0 \leq \gamma \leq 1 + \frac{V(N_I) - V(N_{1P})}{\delta^2} \quad 0 \leq \alpha_V^* \leq 1 \leftrightarrow -\frac{V(N_I) - V(N_{1P})}{\delta^2} \leq \gamma \leq 1$$

il cui soddisfacimento simultaneo è assicurato dall'ipotesi $0 \leq \gamma \leq 1$. Ciò equivale a supporre che, ad esempio, nel tempo le distorsioni relative alle stime preliminari mantengano lo stesso segno e tendano via via a ridursi, oppure che, indipendentemente dalla qualità delle stime preliminari, l'ammontare degli operatori attivi presenti un *trend* decrescente.

4. Stima dei parametri

Il calcolo di tutti i coefficienti incogniti relativi alle espressioni degli stimatori ottimali e dei relativi errori visti nei quattro casi esaminati comporta una stima della distorsione, del parametro di scala e delle varianze degli ammontari preliminari e finali coinvolti nel modello adottato. Pur

rimandando a Rao, Srinath e Quenneville (1989) per maggiori dettagli su tale problema, alcune osservazioni parzialmente euristiche sembrano comunque necessarie.

In primo luogo, la scelta di porre $\alpha = 1$ sia per la stima del livello che della variazione - che renderebbe superflua la stima di qualsivoglia parametro - giustifica la sua applicabilità in quanto permette l'annullamento delle relative distorsioni nel caso in cui $\gamma \cong 1$, situazione probabilmente frequente nell'analisi di flussi animati da un *trend* scarsamente dinamico e da stabilità dei livelli medi degli ammontari in esame. Altrimenti, le tecniche di stima più semplici da seguire nella pratica consisteranno:

1. in un primo caso, nell'assumere la distorsione, la varianza ed il parametro γ al tempo 2 uguali a quelle stimabili al tempo 1 (nel caso di stime trimestrali, si potrebbe fare riferimento ai valori dello stesso trimestre dell'anno precedente).
2. In un secondo caso, ricordando la prima delle ipotesi introdotte nel paragrafo 2, si potrebbe stimare il parametro di scala come rapporto tra gli ammontari (o le medie, supposte disponibili) degli scarti tra stime preliminari e finali ai tempi $(t-1)$ e $(t-2)$; supponendo la stabilità del parametro nel tempo, potrebbe anche essere utilizzata una media aritmetica ponderata dei precedenti rapporti, includendo fino a k coppie di periodi precedenti, con pesi proporzionali alla somma degli ammontari (preliminari, finali od entrambi) relativi ad ogni coppia di periodi. Alternativamente la stima potrebbe basarsi sulla media dei rapporti degli omologhi scarti. Nel primo caso si parlerà dunque di stima basata sul "rapporto tra medie", nel secondo sulla "media dei rapporti". Per la stima della varianza si potrà supporre:

$$V(N_2) - V(N_{2p}) \cong \gamma^2 [V(N_1) - V(N_{1p})] \quad [10]$$

e calcolare la varianza sui valori empirici della [10] (criterio della "varianza corretta"), mentre calcolando la varianza senza la correzione indotta dal parametro di scala si ricorrerà al criterio della "varianza senza correzione".

Inoltre, dovendo optare, nella pratica, tra la scelta ottimale del parametro α in funzione del livello o della variazione, è probabilmente preferibile ottimizzare, in un'ottica congiunturale, la stima della variazione, con una conseguente perdita assoluta di efficienza relativamente alla stima del livello, la cui entità è valutata in Gismondi (1998).

Va infine notato che le procedure descritte nei paragrafi 3 e 4 potrebbero essere sviluppate anche considerando più di un ritardo (ad esempio, le osservazioni preliminari utilizzabili per la

stima potrebbero riferirsi al biennio precedente all'anno da stimare). Da alcune simulazioni preliminari si è constatato come, in generale, in presenza di serie storiche brevi il vantaggio qualitativo associato ad una maggiore complessità delle procedure di stima non sarebbe particolarmente elevato, per cui nel programma – i cui contenuti sono descritti nel paragrafo seguente – sono implementabili i soli stimatori appena descritti, definiti appunto “stimatori con un ritardo”.

5. Le opzioni disponibili nel programma

Il programma, sviluppato in Excel e dotato di alcune macro per visualizzare i vari fogli di lavoro e stampare automaticamente i risultati, è predisposto per il calcolo di tre stimatori:

1. metodo del quoziente;
2. stima ottimale del livello;
3. stima ottimale della variazione.

E' possibile elaborare fino ad un massimo di 120 osservazioni. Per poter calcolare le stime, i diversi metodi richiedono la disponibilità di un numero minimo di osservazioni, pari a 1 per il metodo del quoziente ed a 2 per i metodi ottimali.

Nel programma è predisposto un foglio di lavoro “Input” in cui occorre inserire nell'ordine la variabile temporale (giorni, mesi, anni, ecc.), la variabile che rappresenta l'ammontare preliminare e la variabile che rappresenta l'ammontare finale. Con n riferimenti temporali (cioè n righe oltre alla prima riga che è obbligatoriamente riservata alle etichette dei 3 campi) devono essere inseriti n ammontari preliminari e $(n-1)$ ammontari finali⁸.

I risultati sono disponibili nei fogli “Metodo del quoziente”, “Livello con un ritardo” e “Variazione con un ritardo”. In ognuno dei fogli suddetti sono riportate le seguenti informazioni:

- sigla dello stimatore;
- legenda descrittiva dello stimatore;
- valore medio di α stimato sull'intero periodo (meno che per il metodo del quoziente)

⁸ Ad esempio, si potrebbero inserire la colonna temporale “Anno”, la colonna dell'ammontare preliminare “Gennaio” (se gli ammontari preliminari si riferiscono a tale mese) e la colonna dell'ammontare finale “Totale anno”. Se gli ammontari preliminari vanno dal 1990 al 1999, gli ammontari finali da inserire andranno dal 1990 al 1998, con il 1999 anno da stimare.

- ❑ errore % medio di stima calcolato come media degli errori % di stima (in valore assoluto) relativi ai periodi precedenti a quello effettivo oggetto di stima;
- ❑ rapporto % tra la media quadratica dell'errore e la stima;
- ❑ errore medio di stima della variazione in valore assoluto;
- ❑ valore della stima;
- ❑ valore della variazione stimata rispetto al penultimo periodo della serie (periodo precedente a quello di stima).

Per ognuno degli stimatori calcolati viene anche proposta la versione “Stima corretta con media errori di stima precedenti”. In pratica, se T_0 indica lo stimatore originario e ε indica la media dei rapporti tra valori stimati e valori veri calcolata sugli ultimi periodi 2 e 1 precedenti a quello oggetto di stima (criterio preferibile alla media calcolata su tutti i periodi precedenti), il nuovo stimatore corretto sarà dato da $T_1 = T_0 / \varepsilon$.

Riguardo al metodo del quoziente, esso è espresso nella formula [5].

Lo stimatore ottimale del livello è espresso dalle relazioni [1] e [3], con una MQE data dalla [4].

Lo stimatore ottimale della variazione è espresso dalle relazioni [6] e [8], con una MQE data dalla [9].

Il parametro di scala γ può essere stimato come:

- ❑ “rapporto tra medie”, ossia come il rapporto tra i valori medi che il fenomeno ha assunto nei due ultimi periodi 2 e 1 precedenti a quello di stima;
- ❑ “media di rapporti”, ossia come media dei singoli rapporti tra ammontari relativi agli ultimi tempi 2 e 1 precedenti a quello di stima⁹.

Il parametro δ è invece univocamente stimato come media delle differenze tra valori finali e preliminari relative agli ultimi tempi 2 e 1 precedenti a quello di stima.

Le varianze che compaiono nelle espressioni degli stimatori e dei relativi errori quadratici medi sono stimate calcolando le corrispondenti varianze empiriche, anch'esse riferite ai periodi precedenti a quello di stima. Se le varianze sono “corrette”, significa che sono state moltiplicate per il quadrato del parametro γ , secondo quanto indicato nella formula [10].

⁹ Dovrebbe trattarsi del criterio preferibile in pratica. In pratica si è potuto verificare come, in sede di stima dei parametri γ e δ , considerare solo gli ultimi due periodi completi sia spesso preferibile a considerare tutti i periodi precedenti a quello oggetto di stima.

Tenendo quindi conto del fatto che sia il parametro γ , sia la varianza che compare al membro di sinistra della [10] possono essere stimati in due modi, e che per ogni stimatore calcolato viene anche proposta la versione corretta sulla base degli errori di stima precedenti, il programma calcola in tutto 18 stimatori, di cui 2 relativi al metodo del quoziente, 8 alla stima ottimale del livello ed altrettanti alla stima ottimale della variazione.

La tabella seguente fornisce un prospetto riepilogativo dell'output del programma relativamente agli stimatori riguardanti le procedure ottimali (righe) ed i vari indicatori su di essi disponibili (colonne).

Poiché la stima della media quadratica dell'errore relativamente agli stimatori corretti sulla base degli errori di stima precedenti viene attuata con una procedura "indiretta" – ossia che stima separatamente le varie componenti della MQE senza garantire la coerenza reciproca di tali stima – potrebbero verificarsi dei casi in cui la MQE risulti negativa (e dunque non accettabile), nel qual caso il programma fornisce il risultato "DIV0!" (sotto una radice quadrata compare un elemento negativo).

Infine, si ricorda che lo stimatore della variazione definito dalla [8] è ottimale nel campo degli stimatori *lineari* espressi dalla [1], per cui potrebbero verificarsi casi in cui il metodo del quoziente (che rappresenta evidentemente un caso di stima *non lineare*) possa risultare più efficiente di [8] nella stima di qualche variazione¹⁰.

TABELLA 1 – Riepilogo di stimatori ed indicatori

Legenda	Alfa	Errori %	MQE/Stima%	Errori var.	Stima 1999	Var.% 99/98
Gamma stimato come rapporto tra medie, varianza senza correzione	1,061	7,940	1,624	7,904	46	9,168
Gamma stimato come rapporto tra medie, varianza corretta						
Gamma stimato come media di rapporti, varianza senza correzione						
Gamma stimato come media di rapporti, varianza corretta						
Stima(a1) corretta con media errori di stima precedenti						
Stima(a2) corretta con media errori di stima precedenti						
Stima(b1) corretta con media errori di stima precedenti						
Stima(b2) corretta con media errori di stima precedenti						

¹⁰ Inoltre, si tenga sempre presente che la ottimalità degli stimatori [3] e [8] potrebbe comunque risultare inficiata dalla necessità di stimare alcuni parametri necessari per la loro implementazione empirica.

6. Alcune applicazioni

Le simulazioni empiriche hanno riguardato la stima di indici di variazione medi annui (produzione industriale e vendite al dettaglio) o di ammontari complessivi annui (turismo e trasporti), basate sulla disponibilità delle serie storiche mensili pluriennali così definite:

1. Produzione industriale (da gennaio 1995 a dicembre 1999). Sono state iterate stime separatamente per il comparto minerario, la manifattura e l'energia, mentre l'indice totale è stato stimato sia come media ponderata degli indici parziali che come serie storica autonoma. La disponibilità del dato di gennaio 2000 ha poi consentito di stimare l'indice medio annuo riferito a tale anno.
2. Vendite al dettaglio (da gennaio 1996 a dicembre 1999). Sono state iterate stime separatamente per il comparto alimentare e non alimentare, mentre l'indice totale è stato stimato sia come media ponderata degli indici parziali che come serie storica autonoma. La disponibilità del dato di gennaio 2000 ha poi consentito di stimare l'indice medio annuo riferito a tale anno.
3. Presenze turistiche nelle strutture ricettive (da gennaio 1990 a dicembre 1998). Sono state iterate stime separatamente per le presenze di italiani negli alberghi, italiani nelle strutture complementari (campeggi, villaggi turistici, agriturismi, ostelli ed altre strutture collettive), di stranieri negli alberghi e di stranieri nelle strutture complementari, mentre le presenze totali sono state stimate sia come semplice somma degli ammontari parziali che come serie storica autonoma. La disponibilità dei dati del 1999 fino a settembre ha poi consentito di stimare i totali riferiti a tale anno.
4. Traffico aereo (da gennaio 1995 a dicembre 1998). Sono state iterate stime separatamente per il numero di aerei ed il numero di passeggeri (arrivi internazionali negli aeroporti italiani). La disponibilità dei dati del 1999 fino a settembre ha poi consentito di stimare i totali riferiti a tale anno.
5. Traffico marittimo (da gennaio 1996 a dicembre 1998). Sono state iterate stime separatamente per il numero di navi ed il numero di tonnellate trasportate. La disponibilità dei dati del 1999 fino a giugno ha poi consentito di stimare i totali riferiti a tale anno.
6. Traffico ferroviario (da gennaio 1995 a dicembre 1998). Sono state iterate stime separatamente per il numero di viaggiatori chilometro ed il numero di tonnellate chilometro (espressi nelle tabelle in milioni). La disponibilità dei dati del 1999 fino a marzo ha poi consentito di stimare i totali riferiti a tale anno.

La scelta delle serie storiche è stata ispirata sia dalla necessità di garantirne una certa eterogeneità (proprio per testare l'affidabilità dei vari criteri di stima in condizioni molto diverse), sia dalla effettiva disponibilità dei dati; nel caso delle vendite si è utilizzata la nuova serie storica disponibile a partire da gennaio 1996, mentre le presenze turistiche, pur essendo disponibili per un arco temporale più lungo, sono state soggette negli anni '90 ad alcune modifiche nelle definizioni (specialmente per quello che riguarda le strutture ricettive complementari); l'indagine sui trasporti marittimi è infine quella che ha probabilmente subito, tra il 1997 ed il 1998, le modifiche di maggiore entità.

In tutto sono state dunque condotte 19 stime¹¹ relative ad indici od ammontari annui riferiti al 1998 o al 1999, mentre altre 3 stime dei totali sono state ottenute per media o somma di stime parziali. Va notato come, per applicare il programma di stima nel caso della stima di indici, gli ammontari preliminari e finali sono stati ottenuti per somma dei relativi indici mensili di variazione. In tal modo si è potuto stimare la somma degli indici annui che, divisa per 12 ha consentito di pervenire alla stima dell'indice medio annuo. I 18 stimatori utilizzati sono riepilogati nel prospetto seguente.

PROSPETTO 1 – I 18 stimatori utilizzati per le simulazioni

Sigla	Legenda
1	Metodo del quoziente standard (riproporzionamento)
2	Metodo del quoziente corretto con media errori di stima precedenti
3	Gamma stimato come rapporto tra medie, varianza senza correzione
4	Gamma stimato come rapporto tra medie, varianza corretta
5	Gamma stimato come media di rapporti, varianza senza correzione
6	Gamma stimato come media di rapporti, varianza corretta
7	Stima(a1) corretta con media errori di stima precedenti
8	Stima(a2) corretta con media errori di stima precedenti
9	Stima(b1) corretta con media errori di stima precedenti
10	Stima(b2) corretta con media errori di stima precedenti
11	Gamma stimato come rapporto tra medie, varianza senza correzione
12	Gamma stimato come rapporto tra medie, varianza corretta
13	Gamma stimato come media di rapporti, varianza senza correzione
14	Gamma stimato come media di rapporti, varianza corretta
15	Stima(a1) corretta con media errori di stima precedenti
16	Stima(a2) corretta con media errori di stima precedenti
17	Stima(b1) corretta con media errori di stima precedenti
18	Stima(b2) corretta con media errori di stima precedenti

Più nel dettaglio, la procedura di stima ha operato come segue. Se occorre stimare un indice o un ammontare annuo (1999 per gli indici e 1998 per gli ammontari), tale aggregato annuo è stato stimato 11 volte, considerando come ammontare preliminare quello relativo al solo gennaio, quello

¹¹ 4 per la produzione industriale, 3 per le vendite, 5 per le presenze turistiche e 6 per i trasporti nel complesso.

relativo al bimestre gennaio-febbraio, quello relativo al trimestre gennaio-marzo e così via fino a novembre (quindi all'ammontare relativo ai primi 11 mesi dell'anno).

Per ogni dominio di interesse, ogni stimatore ed ogni ammontare preliminare considerato i risultati dettagliati sono disponibili nelle tabelle in appendice, in cui viene riportato l'errore percentuale di stima (scarto relativo, moltiplicato 100, tra valore stimato e valore vero). In particolare, per ogni stimatore è riportato l'errore medio (basato sugli scarti presi in valore assoluto) relativo alle 11 stime infraannuali (dal solo ammontare preliminare di gennaio a quello relativo ai primi 11 mesi dell'anno), il più alto ed il più basso di tali errori di stima infraannuali, la differenza tra tali valori e gli 11 errori di stima infraannuali, nonché un riepilogo del numero di casi in cui ognuno dei 18 stimatori è risultato ottimale (ossia caratterizzato da un errore di stima inferiore a quello degli altri stimatori). Nel complesso, sebbene si verifichi – come prevedibile – che il ricorso ad ammontari preliminari basati su un numero di mesi più elevato conduca a stime annue più precise, tale evidenza non costituisce una regolarità: ad esempio, nel caso del turismo è solo a partire dal mese di maggio che l'aggiunta di ulteriori mesi osservati comporta una decrescita dell'errore di stima relativo all'ammontare annuo delle presenze, e ciò potrebbe dipendere dalle diverse cadenze delle festività pasquali nel bimestre marzo-aprile.

Una sintesi dei risultati salienti è invece riportata nelle tabelle 2 e 3. Le stime dei totali ottenute per media di stime parziali hanno utilizzato gli stimatori risultati ottimali per ognuno dei sub-domini di riferimento (ad esempio, dalla tabella 2 si deduce che l'indice medio 1999 del totale delle vendite al dettaglio è stato stimato per media ponderata dell'indice alimentare – stimato indifferentemente con gli stimatori ottimali del livello o della variazione senza correzione con gli errori di stima precedenti – e non alimentare, ricavato con lo stimatore ottimale del livello).

L'errore di stima – calcolato nella tabella 2 come media degli 11 errori di stima presi in valore assoluto relativi alle stime basate sul primo mese dell'anno, sui primi 2 mesi, e così via fino a novembre – è dato dallo scarto (percentuale per gli ammontari) tra la stima dell'anno di riferimento e l'ammontare vero relativo a tale anno. Nel complesso, non sembra emergere su tutti uno stimatore certamente preferibile agli altri, sebbene gli stimatori ottimali conducano agli errori di stima più bassi in un maggior numero di casi: il metodo del quoziente è risultato ottimale in 5 casi su 18 (in particolare, per la stima dell'indice della produzione industriale totale), lo stimatore ottimale del livello in 9 casi e lo stimatore ottimale della variazione in 8 casi (in 4 circostanze i 2 stimatori ottimali forniscono stime coincidenti).

Sebbene la correzione con gli errori precedenti conduca spesso ad un significativo incremento del livello medio di errore delle stime (cfr. le tabelle in appendice), in alcuni casi questo accorgimento può condurre ad incrementi di precisione: ciò avviene per la stima dell'indice del

totale della produzione industriale e della produzione manifatturiera, per l'indice delle vendite alimentari e per il numero di viaggiatori chilometro per ferrovia.

Le stime più precise dei "totali" sono state ottenute per le vendite al dettaglio (l'errore di stima è pari allo 0,27%, 0,21% se la stima è diretta, ossia non è ottenuta per media delle stime parziali alimentari e non alimentari), seguite dal numero di passeggeri via aerea (0,44%) e dalla produzione industriale (0,69%, 0,68% se si ricorre alla stima diretta).

Nel caso del turismo, sono risultate più precise le stime relative alle presenze nelle strutture complementari (l'errore è dello 0,38% per gli italiani e dell'1,04% per gli stranieri), in modo almeno in parte sorprendente, data la maggiore instabilità dei flussi turistici in tale tipologia di strutture in confronto con gli alberghi.

Meno precise sono invece risultate le stime delle quantità trasportate via mare (errore del 2,58%) e per ferrovia (5,63%).

Riepilogando, gli stimatori ottimali del livello e della variazione sembrano evidenziare una *performance* superiore rispetto al metodo del quoziente, che sebbene risulti preferibile in diverse circostanze, in altre può condurre ad errori di stima molto elevati.

D'altra parte, il solo esame degli errori medi di stima non consente di evidenziare una caratteristica peculiare del metodo del quoziente, evidente invece dall'esame della tabella 3, che riporta il numero di casi in cui ogni stimatore è risultato ottimale per ogni dominio di interesse (si ricorda che per ogni dominio di interesse i 18 stimatori sono confrontabili 11 volte, tante quante sono le stime effettuate, per cui il totale dei casi considerati è pari a $18 \times 11 = 198$).

In ben 80 casi su 199¹² (il 40,2%) il metodo del quoziente è risultato il migliore (in 58 casi senza ricorso alla correzione con gli errori precedenti), seguito dai 66 casi relativi allo stimatore ottimale del livello (il 33,2%) e dai 53 casi dello stimatore ottimale della variazione (il 26,6%).

Ciò è dovuto al fatto che, in generale, se lo stimatore quoziente è ottimale risulta esserlo indipendentemente dal numero di mesi che contribuiscono a comporre l'ammontare preliminare: ad esempio, ciò si verifica nel trasporto marittimo (numero di navi), nel trasporto ferroviario (tonnellate chilometro) e per la stima delle presenze nelle strutture alberghiere. Con riferimento a quest'ultimo caso è sintomatico sottolineare come gli errori di stima associati alle stime ottimali delle presenze nelle strutture complementari siano, come già detto in precedenza, molto più bassi degli errori medi registrati per le presenze alberghiere, e dato che per le strutture complementari gli stimatori più precisi erano risultati quelli basati sulla stima ottimale del livello si può concludere che, in generale, il metodo del quoziente, pur comportando un errore medio inferiore a quello degli altri stimatori in un numero più elevato di casi, si caratterizza comunque per un livello medio di tale

errore significativamente più alto rispetto a quello registrato dagli stimatori del livello o della variazione nei casi in cui a risultare ottimali siano proprio quest'ultima coppia di stimatori.

In sede previsionale, utilizzando il mese più recente disponibile nelle serie storiche si è cercato di prevedere gli indici medi annui del 2000 sia per le vendite, sia per la produzione industriale, nonché gli ammontari 1999 delle presenze complessive, del numero di passeggeri trasportati su aerei, delle tonnellate trasportate via mare e delle tonnellate chilometro relative al trasporto ferroviario. Per le stime si è fatto ricorso al metodo di stima più efficiente per ciascun dominio di interesse, secondo quanto riportato nella tabella 2¹³.

I valori stimati sono riportati nella penultima colonna della tabella 2, mentre nell'ultima colonna della tabella è riportata la stima della variazione rispetto all'anno precedente.

Ovviamente tali stime hanno un valore indicativo e potranno essere migliorate non appena saranno disponibili i valori relativi a mesi addizionali.

Riferimenti bibliografici

BOLLINO A.C. (1998), "Econometric Interpolation of Regional Time Series", *Statistica applicata*, Vol.10, 1, 35-51, Rocco Curto Editore, Napoli.

GENNARI P., GIOVANNINI E. (1993), "La stima trimestrale dei conti nazionali mediante modelli a parametri variabili", *Quaderni di ricerca*, 5, Istat, Roma.

GISMONDI R. (1996) "Riflessioni sul problema della stima del livello e della variazione tramite l'uso di informazioni preliminari", *Statistica applicata*, Vol.8, 2, 431-453, Rocco Curto editore, Napoli.

GISMONDI R. (1998) "Un metodo per la stima rapida di ammontari tramite risposte preliminari in presenza di errori di misura", *Rassegna dei lavori dell'ISCO*, Anno XV, 1, 73-107, Isco, Roma.

HARRIS C.W. (1963), *Problems in Measuring Change*, University of Wisconsin Press, Madison.

HARVEY A.C. (1984), "A Unified View of Statistical Forecasting Procedures", *Journal of Forecasting*, 3, 245-275.

MARAVALLE M., POLITI M., IAFOLLA P. (1993), "Scelta di indicatori per la stima rapida di un indice provvisorio della produzione industriale", *Quaderni di ricerca*, 6, Istat, Roma.

¹² I casi ottimali sono risultati 199 e non 198 perché in un caso si è verificata la completa coincidenza tra 2 stimatori ottimali.

¹³ Si è stimato direttamente il totale, senza media di parziali. Il metodo del quoziente è stato usato per la produzione industriale ed il trasporto ferroviario, lo stimatore ottimale del livello per il traffico aereo quello della variazione per vendite, turismo e trasporto marittimo.

RAO J.N.K., SRINATH K.P., QUENNEVILLE B. (1989), "Estimation of Level and Change Using Current Preliminary Data", *Panel Surveys*, 457-485, John Wiley & Sons, New York.

APPENDICE - TABELLE