

BANCA D'ITALIA

Temi di discussione

del Servizio Studi

Stime in tempo reale della produzione industriale

di G. Bodo, A. Cividini e L. F. Signorini



Numero 104 - Luglio 1988

BANCA D'ITALIA

Temi di discussione

del Servizio Studi

Stime in tempo reale della produzione industriale

di G. Bodo, A. Cividini e L. F. Signorini

Numero 104 - Luglio 1988

La serie «Temi di discussione» intende promuovere la circolazione, in versione provvisoria, di lavori prodotti all'interno della Banca d'Italia o presentati da economisti esterni nel corso di seminari presso l'Istituto, al fine di suscitare commenti critici e suggerimenti.

I lavori pubblicati nella serie riflettono esclusivamente le opinioni degli autori e non impegnano la responsabilità dell'Istituto.

COMITATO DI REDAZIONE: *IGNAZIO ANGELONI; FRANCESCO FRASCA, GIUSEPPE TULLIO, STEFANO VONA; MARIA ANTONIETTA ORIO (segretaria).*

SOMMARIO

La ricerca presenta un'estensione dello schema per la previsione a breve termine dell'andamento dell'attività produttiva nel settore industriale. Servendosi di modelli statistici basati sui dati giornalieri dei consumi di elettricità, viene presentata una metodologia capace di fornire una valutazione preliminare sulla dinamica produttiva di un certo mese già alla metà dello stesso. L'utilizzo di queste procedure, relativamente poco costose in termini di risorse impiegate, costituisce un altro passo in avanti nella continua ricerca di un quadro congiunturale sempre più tempestivo e completo.

INDICE

Introduzione.....	Pag.	5
1. I metodi univariati.....	"	6
2. L'uso dei dati giornalieri di elettricità.....	"	13
2.1 Considerazioni generali.....	"	13
2.2 Stima e previsioni.....	"	18
3. Combinazioni delle previsioni.....	"	22
4. Conclusioni.....	"	32
Note.....	"	35
Appendice A: Calcolo degli intervalli di confi- denza.....	"	38
Appendice B: La rettifica nella previsione della produzione industriale.....	"	40
Bibliografia.....	"	47

Introduzione (*)

Benchè continui a crescere il peso dei servizi sul reddito e sull'occupazione, il settore industriale mantiene un ruolo fondamentale all'interno dell'economia, presenta le maggiori oscillazioni cicliche tra le varie componenti dell'offerta e tende a trasmetterle a tutto il resto del sistema economico. Da qui l'esigenza di avere informazioni in tempi il più possibile ravvicinati sulla dinamica produttiva del settore. In questa ricerca presentiamo nuovi metodi per stimare in tempo reale l'andamento dell'indice della produzione industriale, nei quali ci si serve per la prima volta dei dati giornalieri sull'energia elettrica. Impiegando questa metodologia a metà mese si può avere un'informazione ragionevolmente soddisfacente sulla dinamica della produzione (1). A fianco di queste procedure impieghiamo anche estrapolazioni a breve termine basate su diversi metodi univariati (ARIMA e Holt-Winters), sia per avere elementi di riferimento, sia per estendere l'orizzonte della previsione. In tal modo è possibile avere stime attendibili del trimestre già alla metà di questo; il guadagno rispetto ai dati ufficiali è di circa un trimestre.

Il piano del lavoro è il seguente: nel successivo paragrafo vengono sinteticamente esposti i risultati ottenuti con l'impiego dei metodi univariati. Nel secondo vengono considerate le problematiche connesse all'uso dei dati giornalieri dell'Enel sui consumi di elettricità. Nel terzo si mostra come previsioni combinate consentano di ottenere ulteriori guadagni di efficienza; seguono quindi le conclusioni. Nell'Appendice B viene discussa in dettaglio la correzione impiegata per tenere conto del diverso numero di giornate lavorative nei vari mesi dell'anno.

1 - I metodi univariati

Naturalmente la prima scelta che è possibile fare per prevedere la dinamica futura delle serie considerate è quella di impiegare metodi univariati, avvalendosi cioè solamente dell'informazione contenuta nei valori passati delle variabili di interesse. Il metodo più comunemente impiegato a tale fine è quello sistematizzato da Box e Jenkins, vale a dire i modelli ARIMA. L'applicazione di questa metodologia alla serie di produzione industriale è già stata discussa in Bodo (1984) e Bodo-Signorini (1985b) e non sarà qui considerata in dettaglio. Sinteticamente, il modello identificato è un semplice schema MA(1,12) con una doppia differenziazione, sequenziale e stagionale, mentre non si è reso necessario trasformare la variabile dipendente. Le consuete analisi in termini di correlogrammi della serie stazionaria, le stime e la verifica dei residui sono riportate in Bodo-Signorini (1985b). Nella figura 1.1 sono riportati i parametri MA(1) ed MA(12) ottenuti nella stima per i vari mesi; la stima di questo modello, come di tutti i successivi, è stata effettuata iterativamente con una procedura di rolling regression in cui le nuove osservazioni eliminano le vecchie allo scopo di approssimare in modo flessibile un modello a coefficienti variabili. In questo modo si è simulata l'applicazione del modello ai fini della generazione di previsioni per il periodo gennaio 1984-aprile 1986 utilizzando solo l'informazione effettivamente disponibile nel periodo di riferimento.

Un secondo metodo comunemente impiegato è quello di Holt-Winters, basato sulla perequazione esponenziale. Abbiamo sperimentato anche questo metodo sia per la sua notevole semplicità d'impiego, sia perchè in certe condizioni dovrebbe risultare particolarmente efficiente. In effetti come è stato mostrato da Granger e Newbold (1977, pp. 163-184) il metodo Holt-Winters può essere ricondotto ad un caso particolare di modello ARIMA del tipo:

$$/1/ \nabla \nabla_{12} x_t = (1 - a_1 B + \sum_{j=2}^{11} a_j B^j - a_{12} B^{12} + a_{13} B^{13}) u_t$$

che è un MA come il modello da noi scelto, ma con più parametri (sebbene non indipendenti fra loro). Nell'applicazione del metodo Holt-Winters abbiamo impiegato una specificazione additiva della stagionalità; per l'inizializzazione si sono adottati i criteri suggeriti da Piccolo e Vitale (1984, p. 646) adattandoli opportunamente alla specificazione additiva. Infine, per stimare i tre parametri del modello (che, come è noto, determinano rispettivamente la velocità di adattamento del livello ciclico, del trend e della stagionalità (2)), si è fatto ricorso ad una funzione di perdita quadratica:

$$/2/ \quad L = \sum_{t=m+h}^n (x_t - f_{t-h,h})^2$$

dove $f_{t-h,h}$ è la previsione ad h periodi di distanza eseguita al tempo $t-h$, n è l'ultimo periodo ed m è sufficientemente grande da limitare gli effetti dovuti all'inizializzazione (nelle nostre stime $m=48$). L'algoritmo per la stima si può basare su una semplice griglia tridimensionale di valori, distanziati in modo opportuno, dei parametri; le terne vengono esaminate una per una e si sceglie quella che minimizza la funzione di perdita. Alternativamente, si può ricorrere alla programmazione non lineare vincolata. Di fatto nel nostro caso i due metodi danno risultati simili e, per semplicità, si è scelto di presentare le stime e le previsioni ottenute con il primo metodo (3).

Nella figura 1.2 sono riportati i parametri ottenuti nelle successive iterazioni della stima; essi variano ma generalmente in misura ridotta nelle diverse stime. Il coefficiente del trend è vicino a zero, il che implica un trend che varia solo molto lentamente. Risulta invece non lontano dall'unità il coefficiente di aggiustamento dei fattori stagionali, mostrando come questi siano variati notevolmente nel corso del periodo considerato.

Le previsioni dei due metodi univariati sono state confrontate con i dati effettivi sulla base dei consueti indici di accostamento (Tav. 1.1 e Fig. 1.3). I risultati appaiono in

genere soddisfacenti. Il coefficiente di correlazione è elevato, mentre contenuti appaiono l'errore medio assoluto e la radice dell'errore quadratico medio che sono compresi tra 2.4 e 3.2; il coefficiente della regressione dei valori effettivi su quelli previsti non è significativamente differente dall'unità (4). Tra i due metodi esiste una moderata superiorità della procedura Holt-Winters rispetto a quella ARIMA per tutti gli orizzonti temporali considerati. Si può anche notare come non vi sia alcuna tendenza al deterioramento della qualità delle previsioni passando dalla previsione a un periodo a quelle a due o tre periodi.

Ai fini di un ulteriore riscontro sull'attendibilità delle stime sono stati anche calcolati, per il modello Holt-Winters, gli intervalli di confidenza all'80 per cento che rappresentano un compromesso fra l'esigenza di non estendere troppo gli intervalli e quella di non avere troppe osservazioni al di fuori di essi. La procedura di calcolo è riportata nell'appendice A. L'ampiezza media risulta assai elevata (11.6 per cento per le previsioni a 1 lead e 12.4 per quelle a 2 lead), ma in entrambi i casi è solo un'osservazione che cade al di fuori degli intervalli. Questa differenza rispetto ai livelli teorici di significatività, oltre che essere attribuita a fattori casuali, può avere diverse interpretazioni: potrebbe anche segnalare problemi relativi alla specificazione del modello e alla normalità degli errori, che però non vengono suggeriti da nessuno dei test impiegati (5). Una spiegazione plausibile consiste nella differenza fra periodo di stima e periodo di previsione: il primo è molto più turbolento e contiene parecchi punti di svolta, il secondo è molto più "liscio". L'ampiezza degli intervalli comunque getta qualche dubbio sull'attendibilità delle previsioni univariate al di là del periodo considerato.

Infine, dai valori mensili sopra considerati si può passare a quelli trimestrali ottenuti per aggregazione dei primi (Tav. 1.2). Come è da attendersi data la maggiore varianza della serie mensile rispetto a quella trimestrale, le previsioni relative a quest'ultima sono nettamente migliori sia in termini dell'U di Theil, sia in termini di errore medio assoluto o

quadratico. Su questa base per il periodo 1984-I 1986-I l'errore medio assoluto che si sarebbe commesso sarebbe stato solo di 1.3 nel caso del metodo Holt-Winters, e di 1.7 nel caso del metodo ARIMA: si tratta di errori molto ridotti anche se la limitatezza del campione considerato (9 osservazioni) induce ad una certa prudenza. Considerando invece il rapporto tra la varianza dei dati stimati e quella dei dati effettivi, le previsioni mensili risultano leggermente preferibili rispetto a quelle trimestrali.

Fig. 1.1

Coefficienti del metodo ARIMA
per i vari periodi di stima

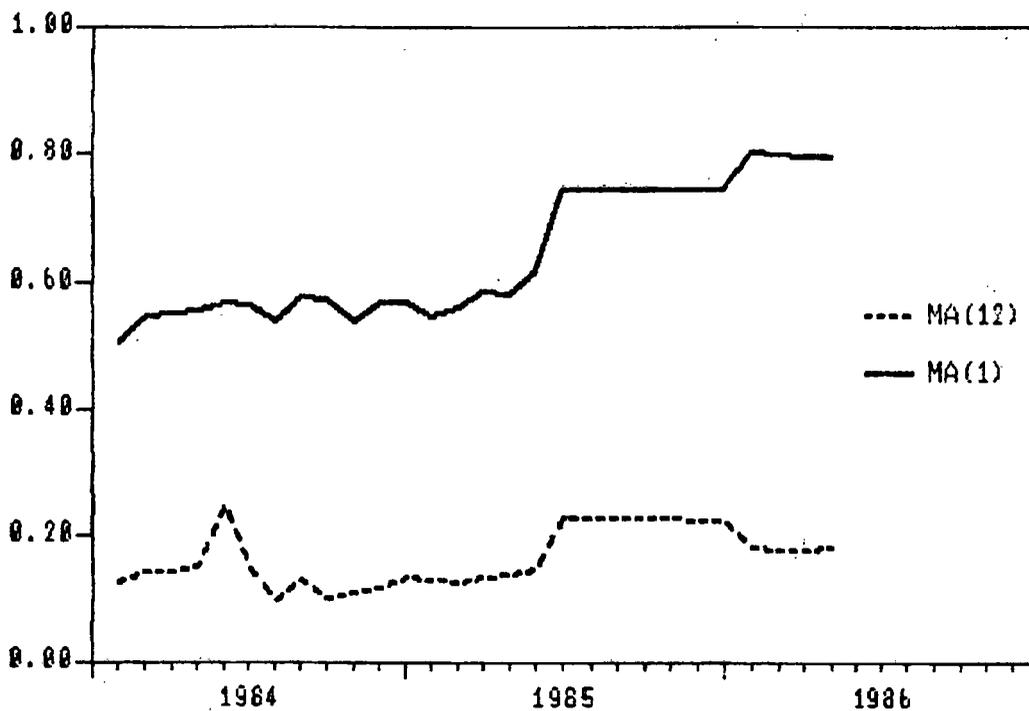


Fig. 1.2

Coefficienti del metodo Holt-Winters
per i vari periodi di stima

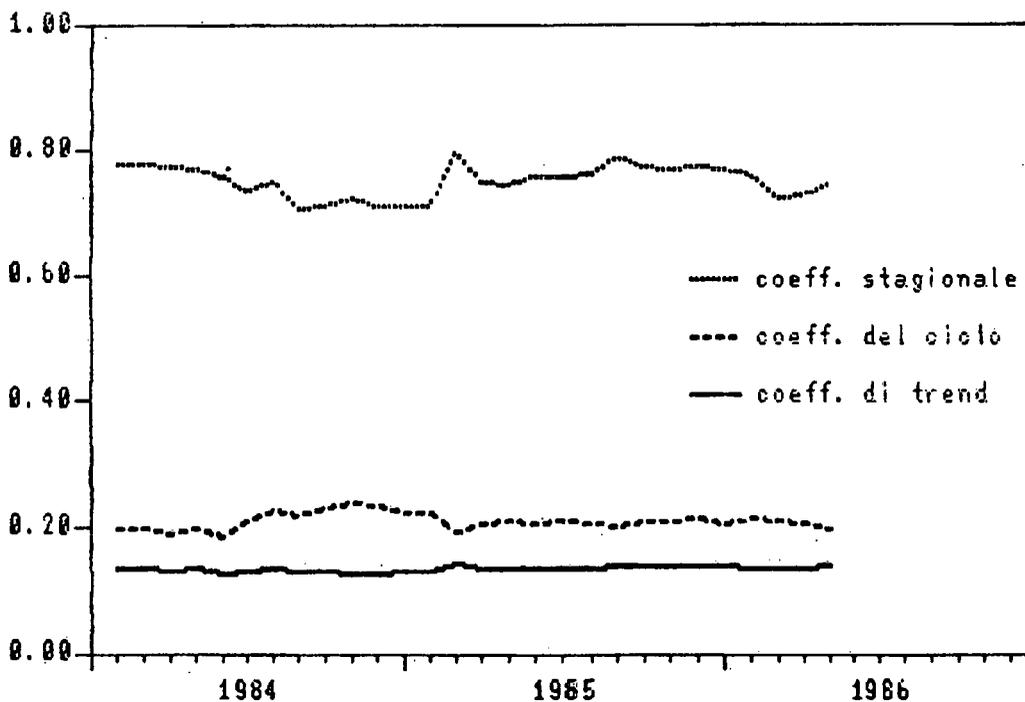
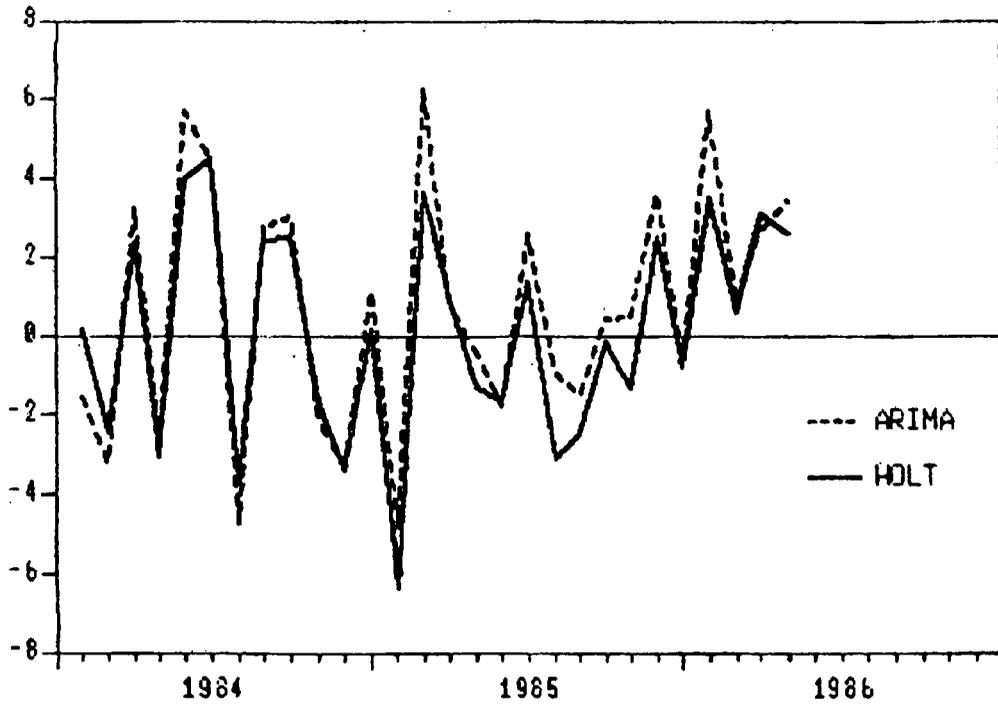


Fig. 1.3

Errori di previsione con i diversi metodi
(1 lead)



Confronto tra previsioni e realizzazioni
con i metodi univariati (8401-8604)

TAV. 1.1

	1 Lead		2 Lead		3 Lead	
	HW	ARIMA	HW	ARIMA	HW	ARIMA
r	.98	.98	.98	.98	.98	.98
U	.0278	.0321	.0278	.0311	.0281	.0322
b	.998	.990	1.009	1.020	.995	.994
ME	.08	.67	.16	.82	.29	1.07
MAE	2.35	2.68	2.36	2.42	2.31	2.54
RMSE	2.74	3.17	2.72	3.07	2.77	3.17
F-prob	98.9	53.8	92.6	33.6	85.8	20.4
PM	.034	.043	.033	.039	.034	.040

Confronto tra previsioni e realizzazioni
con i metodi univariati (8401-8604) - dati trimestrali

TAV. 1.2

	Holt-Winters	ARIMA
r	.98	.96
U	.0164	.0219
b	1.055	1.037
ME	-.28	.29
MAE	1.26	1.71
RMSE	1.60	2.13
PM	.045	.081

Legenda: r, coefficiente di correlazione, U coefficiente di Theil (seconda definizione), b coefficiente angolare della regressione dei valori effettivi su quelli previsti, RMSE radice dell'errore quadratico medio, MAE errore medio assoluto, ME errore medio, F-prob livello di significatività del test-F sulla restrizione b uguale ad 1 e costante uguale a zero, PM rapporto tra la varianza dell'errore e la varianza dei valori effettivi.

2 - L'uso dei dati giornalieri di elettricità.

2.1 Considerazioni generali

L'impiego dei dati di elettricità è ormai esteso nell'analisi congiunturale di breve periodo (6). Sebbene in forma diversa (uso dei consumi elettrici totali o per circoscrizione, differenti specificazioni dinamiche, diverse variabili considerate in aggiunta ai consumi di elettricità) tutti questi modelli si servono di dati mensili dei consumi elettrici, e quindi permettono di stimare l'indice della produzione industriale solo all'inizio del mese successivo. Noi abbiamo provato a utilizzare anche i dati sull'immissione giornaliera di energia elettrica nella rete dell'ENEL. Naturalmente, volendo stime mensili i dati giornalieri interessano solo per l'informazione che forniscono sull'andamento del mese prima che questo sia finito, diciamo dopo g giorni; il valore di g deve essere scelto in modo da ottenere stime da un lato sufficientemente anticipate, dall'altro sufficientemente attendibili. Sulla base di considerazioni a priori e di alcune prove preliminari, abbiamo deciso di considerare i primi 14 giorni del mese. Poco dopo la metà del mese, tenendo conto dei tempi di raccolta e di elaborazione dei dati, sarebbe possibile avere una stima dell'indice corrente.

Per realizzare il modello previsivo occorre specificare: (a) la forma della funzione che lega elettricità e produzione; (b) il trattamento dei giorni non lavorativi e in generale delle differenze di calendario; (c) il modo di trasformare i dati quattordicinali in mensili.

Il primo punto non può essere discusso ora in modo esteso; si rinvia a Bodo-Signorini (1985b) dove si tratta anche il problema dell'impiego dei soli dati ENEL (che escludono autoproduttori e aziende municipalizzate). Il modello ivi sperimentato con i dati mensili viene qui sostanzialmente riproposto e adattato ai dati inframensili. Le caratteristiche principali del modello sono le seguenti:

1. la relazione ipotizzata è fondamentalmente lineare;
2. trend e stagionalità vengono trattati con dummies appropriate e termini ritardati della variabile dipendente;
3. gli "effetti di mix" della produzione industriale (variazioni della quota dei consumi delle industrie ad alta intensità elettrica) vengono approssimati utilizzando dati per ripartizioni territoriali ("compartimenti");
4. dei consumi non industriali di elettricità si tiene conto (tra l'altro) per mezzo della temperatura introdotta in forma quadratica.

La questione della rettifica per le variazioni di calendario è affrontata nell'Appendice B. Sinteticamente, nella relazione stimata la produzione industriale viene rettificata dividendola per il numero delle giornate lavorative del mese (produzione media giornaliera), ma tenendo anche conto delle lavorazioni a ciclo continuo, che pesano per il 15% circa sull'indice. L'elettricità viene rettificata in modo simile, ma attribuendo un peso maggiore ai giorni non lavorativi sia per la maggiore intensità elettrica delle lavorazioni a ciclo continuo sia (limitatamente al secondo dei due metodi di stima proposti qui appresso) per l'influenza dei consumi domestici e non industriali in genere.

Quanto all'ultimo punto, si possono seguire due strade: la prima ("metodo diretto") consiste nello stimare la relazione tra elettricità e produzione con i dati dei primi g giorni anziché il totale mensile; la seconda ("metodo indiretto") nello stimare l'elettricità mensile con i dati dei primi g giorni e poi, sulla base di questa, la produzione con il metodo usuale. Esamineremo entrambi i metodi in maggior dettaglio.

Metodo diretto. Il modello elettricità-produzione mensile è il seguente (7):

$$\text{/3/ } Y = (Y^M/l_Y) = A + B(E^M/l_e) + CT + F(L)(Y_M/l_Y) + \text{dummies} \\ \text{stagionali} + u$$

dove:

Y^M , E^M sono rispettivamente la produzione e l'elettricità mensili grezze;

T rappresenta la/le variabili (temperatura, ...) che influenzano i consumi extraindustriali;

l_Y è il coefficiente di rettifica ponderato della produzione industriale;

l_e è il coefficiente di rettifica ponderato dei consumi elettrici industriali: si distingue dal precedente perchè le lavorazioni continue pesano il 45% invece del 15% circa;

$F(L)$ è un polinomio nell'operatore di ritardo L;

u è un errore casuale.

Il suffisso temporale pertinente alle variabili Y^M , E^M , T, ai coefficienti di rettifica e al residuo è sottinteso; A, B, C, i coefficienti delle dummies e di F sono parametri da stimare. La produzione media giornaliera dei primi g giorni del mese, $Y' = Y^{M'}/l'_Y$ può essere stimata sulla base dei corrispondenti dati elettrici in modo formalmente identico alla /3/, purchè ovviamente vengano utilizzati i coefficienti di rettifica appropriati, cioè quelli riferiti ai primi g giorni (che in avanti denotiamo con un apice). Y' , tuttavia, non interessa per sé ma per le informazioni che fornisce sulla produzione media del mese, Y. Il valore atteso di Y, noto Y' , è

$$\text{/4/ } \text{EXP}(Y | Y') = Y' + \text{EXP}(\dot{Y} - Y')$$

L'ipotesi più semplice è che l'attività produttiva sia costante nel mese. Se è così l'ultimo termine della /4/ è nullo: la produzione totale (stimata) del mese, \dot{Y}^M , è semplicemente uguale a quella dei primi giorni, $\dot{Y}^{M'}$, moltiplicata per il

rapporto tra i rispettivi coefficienti di rettifica:

$$/5/ \quad \hat{Y}^M = \hat{Y}^{M'} * (1_{y} / 1_{y'})$$

Ma questa ipotesi non è realistica. A parte oscillazioni casuali che per definizione non incidono sul valore atteso di $Y-Y'$, i possibili motivi di discrepanza fra le due produzioni medie sono i seguenti:

1. se l'attività produttiva ha una componente tendenziale, la produzione della seconda metà del mese sarà sistematicamente superiore (o inferiore) a quella della prima metà;
2. la componente ciclica ha un effetto analogo, ma variabile nel tempo a seconda della fase congiunturale attraversata;
3. la distribuzione prevalente di eventuali ferie e "ponti" (che, nella misura in cui interessano giorni feriali, non vengono presi in considerazione dalla rettifica), e in generale delle variazioni stagionali sistematiche, non è necessariamente omogenea nel corso del mese: questo problema riguarda soprattutto determinati mesi dell'anno, quali gennaio, agosto, novembre, dicembre e il mese in cui cade la Pasqua.

Il punto 1 incide sulla costante; il punto 2 principalmente sui parametri dei termini ritardati di Y ; il punto 3 sulle dummies stagionali (in più suggerisce l'opportunità di inserire nella relazione la dummy "Pasqua nei primi g giorni del mese"/"Pasqua nei restanti giorni del mese"). Ne segue che la relazione produzione mensile \leftrightarrow elettricità inframensile deve essere stimata in modo indipendente rispetto a quella basata solo sui dati mensili; i parametri saranno in generale diversi. La stima deve avvenire su dati rettificati come nella /3/, usando

però i coefficienti inframensili. La produzione grezza mensile può poi essere ricavata dalla rettificata mensile nel modo usuale.

Metodo indiretto. Per stimare nel modo più accurato possibile l'elettricità mensile (primo stadio del metodo) devono ovviamente essere usati i coefficienti di rettifica propri dell'elettricità totale; si noti che invece con il metodo diretto l'elettricità viene rettificata con i coefficienti relativi ai soli consumi elettrici industriali, per i motivi esposti nell'Appendice B. Poichè le cause di variazioni sistematiche inframensili elencate sopra per la produzione valgono mutatis mutandis anche per l'elettricità, la stima deve essere compiuta sulla base di una regressione della seguente forma:

$$\text{/6/} \quad (E^M/n_e) = a + b(E^{M'}/n'_e) + c(L)(E_M/n_e) + \text{dummies} + \text{errore,}$$

dove $c(L)$ è un polinomio di grado opportuno, n_e è il coefficiente di rettifica dell'elettricità totale e l'apice denota ancora una volta grandezze riferite ai primi g giorni del mese.

Nel secondo stadio del metodo si deve applicare la /3/ all'elettricità stimata con la /6/; i criteri per la rettifica a questo punto sono quelli validi per i dati mensili. Se la stima /6/ è compiuta in modo corretto si devono adoperare gli stessi parametri del modello mensile, cioè quelli stimati con i dati elettrici definitivi, perchè la stima di un modello separato basato sui dati elettrici provvisori soffrirebbe dei noti inconvenienti propri dei modelli di regressione con errori di misura delle variabili indipendenti.

2.2 Stima e previsioni

La relazione /3/ è stata stimata sullo stesso periodo dei metodi univariati; analogamente si sono generate serie di previsioni per il periodo 8401-8604. Trattandosi di stime in tempo reale, queste serie si rendono disponibili con due mesi di anticipo sul dato ufficiale; sono quindi comparabili alle previsioni con lead 2 dei metodi univariati. Le stime basate sull'elettricità di tutto il mese sono comparabili a previsioni con lead 1. Anche qui è stata usata una rolling regression (cfr. paragrafo 1).

Per mettere in evidenza l'utilità della rettifica le stime sono state effettuate sia con i dati grezzi sia con quelli rettificati. Per ragioni di spazio le stime non vengono riportate; nelle loro linee generali sono molto simili a quelle presentate in Bodo-Signorini (1985b). Elettricità, temperatura e dummies stagionali (per la maggior parte) sono significative così come la dummy per la Pasqua; non vi sono segni di autocorrelazione né di instabilità strutturale. I risultati, in termini di capacità previsiva, sono riportati nella tabella 2.1.

Dal confronto con i metodi univariati risulta evidente che i dati quattordicinali grezzi dell'elettricità (metodo diretto) forniscono previsioni non solo peggiori di quelli mensili, ma addirittura inferiori ai metodi univariati. Il giudizio è completamente diverso per le previsioni basate sui dati rettificati; rispetto ai dati grezzi l'errore medio assoluto è di 2.31 contro 2.95 e la radice dell'errore quadratico medio è di 2.79 contro 3.46. Nel confronto con i metodi univariati, le previsioni basate sui dati giornalieri rettificati secondo il metodo diretto presentano un chiaro vantaggio rispetto all'Arima a 2 lead, mentre rispetto al metodo Holt-Winters, sempre a 2 lead, hanno un valore più elevato per l'U di Theil e la radice dell'errore quadratico medio (2.79 contro 2.72), ma un valore meno elevato per l'errore medio assoluto (2.31 contro 2.36).

Le conclusioni che si possono trarre sulla base della procedura indiretta di stima sono molto simili a quelle già ottenute con la procedura diretta. Anche in quest'ultimo caso l'uso

dei dati giornalieri, se opportunamente rettificati sembra nettamente superiore al metodo ARIMA; questa volta anche rispetto all'Holt-Winters vi è un lieve miglioramento in tutti gli indici.

Per le previsioni ENEL sono stati poi calcolati gli intervalli di confidenza all'80 per cento. Essi risultano considerevolmente più ristretti di quelli relativi ai metodi univariati, sia per il modello mensile, sia per quello quattordicinale. Per il primo l'ampiezza media è del 7.2 per cento; solo due osservazioni cadono al di fuori, con un livello di significatività ex-post del 7 per cento rispetto al 20 per cento teorico. Valgono le stesse considerazioni fatte a questo proposito sui modelli univariati. Per quanto riguarda il modello quattordicinale, gli intervalli sono di ampiezza maggiore (9 per cento), ma sempre molto più ristretti di quelli relativi alle previsioni univariate; il numero delle osservazioni che cadono al di fuori (cinque, pari al 18 per cento) si avvicina qui a quello atteso.

E' poi possibile utilizzare le stime basate sull'elettricità per prevedere la dinamica trimestrale con i dati disponibili a metà trimestre. La stima relativa al primo mese del trimestre (previsione a un passo di distanza) è quella ENEL mensile. Per le previsioni a 2 passi si usa l'Enel quattordicinale (è stata utilizzata la metodologia di stima diretta, ma la procedura indiretta dà risultati molto simili). Per completare il trimestre si deve usare uno dei due metodi univariati. Nella tavola 2.2 sono riportati i valori dei soliti indicatori di adattamento delle previsioni Enel integrate per il terzo mese con i dati derivanti sia dal metodo ARIMA, sia da quello Holt-Winters. I risultati ottenuti non variano sensibilmente tra l'integrazione con la procedura Holt-Winters e l'integrazione con il metodo ARIMA. Rispetto a quanto ottenuto con le previsioni univariate si nota una netta superiorità delle stime Enel trimestrali. E' inoltre da rilevare che, a differenza dei metodi univariati, le previsioni trimestrali sembrano più attendibili di quelle mensili, anche tenendo conto della minore varianza delle serie trimestrali.

In conclusione, le previsioni basate sull'elettricità,

oltre ad essere superiori in termini di adattamento a quelle univariate, sono anche nettamente più affidabili - come dimostra l'analisi degli intervalli di confidenza - e consentono di migliorare le stime trimestrali. Questa affermazione può essere rafforzata in base alla considerazione che, nell'arco di tempo considerato per la previsione (8401-8604), non vi sono stati ampi movimenti del ciclo, nè passaggi da fasi di espansione a fasi di recessione o viceversa; tutti fenomeni che in una prospettiva più ampia, almeno a priori, sembrano dover favorire nella previsione l'impiego dei consumi elettrici, piuttosto che dei metodi univariati (8).

TAV. 2.1

Confronto tra realizzazioni e previsioni basate sui
consumi di elettricità (8401-8604)

Indicatori	Enel mensile	Enel 14 grezzo stima diretta	Enel 14 rettificato stima diretta	Enel 14 rettificato stima indiretta
r	.99	.98	.98	.99
U	.0213	.0350	.0283	.0270
b	.998	.980	.986	.987
ME	.19	1.59	.77	.89
MAE	1.73	2.95	2.31	2.26
RMSE	2.10	3.46	2.79	2.66
F. prob.	90.0	4.3	33.6	19.7
PM	.020	.043	.032	.028

TAV. 2.2

Confronto tra previsioni e realizzazioni con le
stime Enel integrate dai metodi univariati
(8401-8601) - Dati trimestrali

	Enel con ARIMA	Enel con Holt-Winters
r	.99	1.00
U	.0159	.0126
b	1.002	1.012
ME	1.23	1.01
MAE	1.34	1.07
RMSE	1.55	1.22
PM	.016	.008

Legenda: vedi TAVV. 1.1-1.2

3 - Combinazione delle previsioni

La combinazione tra i diversi metodi è stata oggetto recentemente di notevole attenzione e numerosi studi teorici ed empirici hanno trattato di questo argomento (9). Considerando per semplicità il caso di due stime non distorte \hat{y}_1 e \hat{y}_2 della variabile y , gli errori di previsione sono definiti come:

$$/7 / \quad e^{(j)} = y - \hat{y}_j ; \quad E (e^{(j)2}) = \sigma_j^2 \quad \text{per } j = 1,2$$

Poichè le previsioni non sono distorte, e quindi il valore atteso di $e^{(j)}$ è pari a zero, una qualunque media ponderata delle due previsioni con pesi che sommano all'unità:

$$/8 / \quad \hat{y} \equiv a \hat{y}_1 + (1 - a)\hat{y}_2$$

sarà anch'essa una previsione non distorta. La varianza dell'errore della previsione combinata sarà pari a:

$$/9 / \quad \sigma_y^2 = a^2 \sigma_1^2 + (1-a)^2 \sigma_2^2 + 2a(1-a)\rho\sigma_1\sigma_2$$

dove $\rho \sigma_1 \sigma_2$ è la covarianza tra i due errori di previsione

La varianza viene quindi minimizzata per un valore di a pari a:

$$/10/ \quad a = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

Tuttavia, la formula /10/ non è utilizzabile immediatamente poichè in pratica non si conoscono i valori di σ_1 , σ_2 e ρ . Disponendo comunque delle informazioni necessarie è possibile, per ogni periodo t , stimare il peso ottimale a_t sulla base dei dati campionari. La stima di massima verosimiglianza è quella data dei minimi quadrati ordinari, cioè:

$$/11/ \quad a_t = \frac{\sum_{t=n-v}^{n-1} (e_t^{(2)2} - e_t^{(1)} e_t^{(2)})}{\sum_{t=n-v}^{n-1} (e_t^{(1)2} + e_t^{(2)2} - 2e_t^{(1)} e_t^{(2)})}$$

dove $v=n-1$ se si desidera compiere la stima sull'intero campione, oppure viene opportunamente scelto per avere pesi flessibili (rolling regression). Tuttavia in quest'ultimo caso (che appare preferibile per la natura della relazione considerata) la stima /11/, benchè teoricamente ottimale, non è in pratica utile perchè i coefficienti di correlazione non sono stimati in maniera soddisfacente nei piccoli campioni. Per questi motivi in letteratura si impiegano delle approssimazioni.

Nel nostro caso abbiamo basato le combinazioni delle previsioni su due metodi che, a parte la presenza della variabile dipendente sfasata, dovrebbero essere indipendenti, l'Holt-Winters e le stime sui dati di elettricità; si sono accantonati i risultati della procedura ARIMA, molto correlati a quelli ottenuti con l'Holt-Winters e meno precisi di questi ultimi. Sulla base di precedenti esperienze in materia, abbiamo

sperimentato tre diverse definizioni del peso a_t :

$$/12/ \quad a_{1,t} = \left[\frac{\sum_{s=t-\tau}^{t-1} e_s^{(1)2}}{\sum_{i=1}^2 \sum_{s=t-\tau}^{t-1} e_s^{(i)2}} \right]^{-1}$$

$$/13/ \quad a_{2,t} = \left[\frac{\sum_{s=1}^{t-1} \alpha^s e_s^{(1)2}}{\sum_{i=1}^2 \sum_{s=1}^{t-1} \alpha^s e_s^{(i)2}} \right]^{-1}$$

$$/14/ \quad a_{3,t} = \beta a_{3,t-1} + (1 - \beta) a_{1,t}$$

In tutte le formule i pesi da attribuire a ciascun metodo fanno riferimento agli errori percentuali di previsione commessi nei periodi precedenti (10). Nella formula /12/ si fa riferimento ad un numero costante τ di errori; l'ultima osservazione elimina la prima. Nella formula /13/ si utilizzano tutti gli errori osservati, compresi i più lontani nel tempo, ma con un peso decrescente in ragione geometrica α . La formula /14/ è un exponential smoothing della /12/ con parametro di attenuazione β . In tutte le formulazioni impiegate si è ignorata la correlazione degli errori poichè i risultati ottenuti in letteratura con procedure che tengono conto della correlazione sono stati largamente inferiori a quelli che la ignorano.

Sulla base dell'esperienza, del numero delle osservazioni e di prove preliminari si sono attribuiti i seguenti valori ai parametri α , β e τ :

$$\alpha = 1.5 \text{ o } 2; \quad \beta = 0.5 \text{ o } 0.7 \quad \tau = 3 \text{ o } 6$$

In aggiunta alle formule sopra esposte, abbiamo determinato i pesi utilizzando una regressione ausiliaria tra il dato storico e le diverse previsioni; poichè, come nel caso teorico, i risultati migliori sono stati ricavati con una

regressione in cui la costante era pari a zero e la somma dei coefficienti delle singole previsioni era pari all'unità, nelle tabelle di confronto con i dati effettivi sono riportate solo le stime ottenute con le restrizioni.

Abbiamo poi impiegato anche la metodologia del filtro di Kalman sempre nel tentativo di avere una procedura di aggiornamento efficiente delle previsioni combinate; in quest'ultimo caso, però, l'adattamento ai dati osservati non è risultato migliore di quello avuto dalle formule molto più semplici prima esposte; perciò dato il suo costo elevato di elaborazione, tale soluzione è stata scartata.

Infine come caso particolare è stata anche esaminata la combinazione delle previsioni per media aritmetica.

Nelle tavole 3.1 - 3.3 sono riportati i confronti tra dati previsti e realizzazioni sempre per il periodo 8401-8604, per le combinazioni considerate che sono:

- a) Holt-Winters a 1 lead, Enel mensile
- b) Holt-Winters a 2 lead, Enel quattordicinale (metodo diretto)

Dall'esame degli indicatori riportati si può notare come con tutte le combinazioni proposte vi sia un netto miglioramento della qualità delle previsioni ad 1 passo sia in termini di radice dell'errore quadratico medio, sia in termini di errore medio assoluto; rispetto alla migliore singola previsione si ha una riduzione massima dell'11 per cento nella radice dell'errore quadratico medio e del 18 per cento nell'errore medio assoluto. Le stesse conclusioni valgono anche per le combinazioni a 2 periodi: l'errore medio assoluto e la radice dell'errore quadratico medio sono entrambi sempre inferiori a quelli del migliore metodo singolo; le riduzioni massime in questo caso sono pari, rispettivamente, all'8 e al 10 per cento. Il coefficiente di correlazione tra gli errori medi assoluti e le radici dell'errore quadratico medio dei diversi metodi a 1 e 2 lead è molto vicino a zero. Non emerge chiaramente la superiorità di un metodo di ponderazione rispetto a un altro; è tuttavia da notare che la semplice media aritmetica presenta la minore distanza dai

dati effettivi a 1 e a 2 periodi; si tratta di un risultato comune in letteratura e conferma quanto era stato ottenuto in Bodo-Signorini (1985b). Tra le altre combinazioni sembrano fornire buoni risultati anche la formula /12/ e la regressione con i minimi quadrati ordinari. Nella figura 3.1 è mostrato come variano i pesi delle stime ottenute con le previsioni sui dati di elettricità a 1 lead secondo i tre migliori metodi di combinazione.

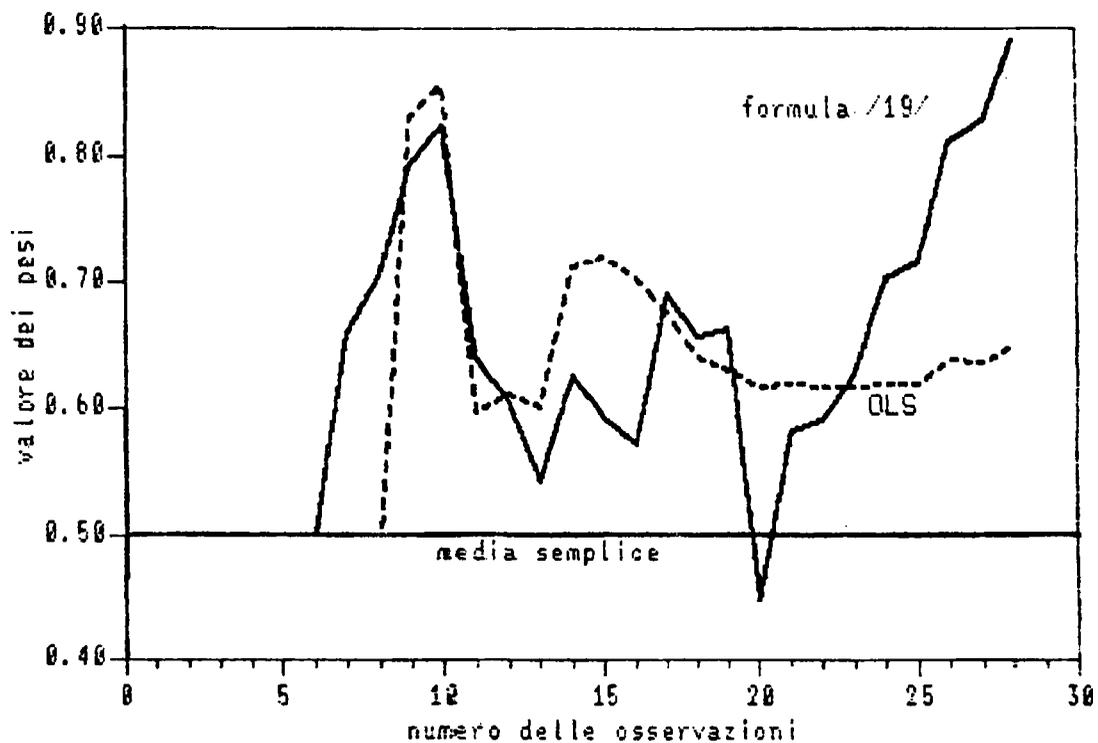
E' poi anche possibile passare dalle previsioni ottenute ai dati trimestrali che quindi costituiscono un'aggregazione di quanto ottenuto con le varie formule, utilizzando il dato Holt-Winters per la previsione a 3 lead. Anche in questo caso tutte le procedure di combinazione presentano un accostamento più soddisfacente rispetto ai migliori risultati derivanti dai singoli metodi; le maggiori riduzioni per l'errore quadratico medio e per l'errore medio assoluto sono pari, rispettivamente, al 22 e al 39 per cento, valori superiori a quelli registrati per le combinazioni a 1 e 2 lead. Sempre rispetto a queste ultime l'analisi dei dati trimestrali mostra una notevole convergenza nei valori degli errori commessi da tutti i vari metodi impiegati.

Infine, come per le stime basate sui dati di elettricità, si sono calcolati gli intervalli di confidenza anche per le combinazioni delle previsioni. Tale procedimento è infatti possibile poichè gli errori di previsione del metodo Holt-Winters e del metodo Enel sono distribuiti normalmente, mentre la varianza della previsione combinata è data dalla formula /9/. Poichè teoricamente i due metodi sono ottenuti in maniera indipendente, abbiamo posto uguale a zero il coefficiente di correlazione tra i due errori di previsione (11). Data la varianza degli errori delle previsioni combinate abbiamo calcolato gli intervalli di confidenza, sempre al livello di significatività del 20 per cento. Nella tabella 3.4 sono riportati il numero di osservazioni che cadono al di fuori degli intervalli di confidenza, che passano da un'ampiezza del 7.2 per cento della previsione Enel a una minima del 6.4 per cento nelle combinazioni. A 1 periodo di distanza tutte le formule di combina-

zione riducono l'ampiezza degli intervalli, pur se, in genere, aumenta il numero di osservazioni che cadono al di fuori dei limiti fissati; il limite teorico del 20 per cento è comunque sempre largamente soddisfatto. Per quanto riguarda le previsioni a 2 periodi, i risultati sono ancora più favorevoli. La riduzione dell'ampiezza degli intervalli di confidenza è molto maggiore di quella riscontrata per le previsioni a 1 passo; è una conclusione che di nuovo vale per tutte le combinazioni proposte. Inoltre, a differenza della situazione precedente, si ha contemporaneamente una riduzione delle bande di confidenza e una diminuzione nel numero delle osservazioni che cadono al di fuori, un risultato particolarmente interessante.

Fig. 3.1

Andamento del peso della previsione Enel
per la combinazione a 1 passo secondo
varie formule



TAV. 3.1

Confronto tra previsioni e realizzazioni:
 combinazioni dei diversi metodi (8401-8604) - Lead 1

FORMULE	RMSE	MAE	ME
Media aritmetica	1.86	1.41	.13
OLS ristretto	1.97	1.51	.24
/12/con $\tau=3$	1.98	1.50	-.03
/12/con $\tau=6$	1.92	1.46	.15
/13/con $\alpha=1.5$	1.92	1.47	.09
/13/con $\alpha=2.0$	1.93	1.48	.03
/14/con $\beta=0.5$	2.00	1.53	.01
/14/con $\beta=0.7$	2.02	1.55	.02

TAV. 3.2

Confronto tra previsioni e realizzazioni:
 combinazioni dei diversi metodi (8401-8604) - Lead 2

FORMULE	RMSE	MAE	ME
Media aritmetica	2.51	2.08	.46
OLS ristretto	2.51	2.13	.46
/12/con $\tau=3$	2.61	2.24	.54
/12/con $\tau=6$	2.53	2.13	.46
/13/con $\alpha=1.5$	2.58	2.18	.50
/13/con $\alpha=2.0$	2.62	2.25	.55
/14/con $\beta=0.5$	2.53	2.13	.50
/14/con $\beta=0.7$	2.51	2.10	.48

Legenda: RMSE radice dell'errore quadratico medio, MAE errore medio assoluto, ME errore medio.

Confronto tra previsioni e realizzazioni:
 combinazioni dei diversi metodi
 (8401-8604) - Dati trimestrali

FORMULE	RMSE	MAE	ME
Media aritmetica	1.01	.68	.37
OLS ristretto	.99	.69	.51
/12/con $\tau=3$.98	.74	.43
/12/con $\tau=6$.95	.65	.48
/13/con $\alpha=1.5$.96	.67	.48
/13/con $\alpha=2.0$.97	.68	.47
/14/con $\beta=0.5$	1.02	.80	.42
/14/con $\beta=0.7$	1.03	.82	.41

Legenda: RMSE radice dell'errore quadratico medio, MAE errore medio assoluto, ME errore medio.

Intervalli di confidenza e combinazioni
delle previsioni (*)

	1 periodo		2 periodi	
	A	B	A	B
Enel	7.2	2	9.0	5
Media aritm.	6.8	2	7.6	3
OLS ristretto	6.4	4	7.8	3
/12/con $\tau=3$	6.8	3	7.8	4
/12/con $\tau=6$	6.4	4	7.7	3
/13/con $\alpha=1.5$	6.4	4	7.7	4
/13/con $\alpha=2.0$	6.6	4	7.8	4
/14/con $\beta=0.5$	6.7	4	7.7	4
/14/con $\beta=0.7$	6.8	3	7.8	3

! A - Ampiezza media dell'intervallo di confidenza in percentuale!
! del dato storico.
! B - Numero di casi al di fuori dell'intervallo.
!

(*) Intervalli all'80 per cento

Conclusioni

Uno dei principali requisiti per un'efficace analisi congiunturale è dato dalla rapida disponibilità di indicatori appropriati. Nel caso dell'industria, settore chiave all'interno del quadro reale per comprendere l'evoluzione ciclica dell'economia, l'indicatore di riferimento, la produzione industriale, viene reso noto con un ritardo di circa due mesi dal termine del mese considerato. Al fine di eliminare tale ritardo abbiamo proposto una metodologia che consente di stimare in tempo reale l'indice della produzione industriale. La variabile chiave per effettuare queste previsioni è il consumo di elettricità, disponibile, oltre che a frequenza mensile, anche giorno per giorno. Questa serie non è un leading indicator in senso stretto; ciò che la rende particolarmente interessante è la possibilità di disporne praticamente in tempo reale. I risultati ottenuti con i consumi elettrici mensili, con orizzonte previsivo ad 1 mese, sono ormai noti; invece non erano ancora stati impiegati i dati sui consumi elettrici giornalieri. Sulla base di considerazioni teoriche e di prime analisi empiriche, abbiamo deciso di utilizzare i dati delle prime due settimane del mese in corso per ottenere stime dell'indice di produzione dello stesso periodo. I dati dei consumi elettrici possono essere inseriti nella relazione tra produzione industriale e consumi di elettricità, ristimandola ex novo (stima diretta). In alternativa, i dati delle prime due settimane possono essere impiegati per ottenere i consumi di elettricità mensili e da questi, sulla base della relazione già stimata, l'indice di produzione industriale (stima indiretta). In entrambi i casi è cruciale tenere conto del problema delle giornate lavorative a frequenza inframensile; l'inserimento dei consumi elettrici tout-court senza alcuna correzione porta infatti a risultati peggiori di quelli ottenibili con l'impiego di semplici modelli univariati. Se invece i consumi di elettricità dei primi 14 giorni vengono trattati opportunamente per tenere conto del numero di giornate lavorative, si arriva sia con la stima diretta, sia con quella

indiretta a previsioni complessivamente migliori di quelle derivabili dai metodi univariati e non troppo lontane da quelle basate sui consumi dell'intero mese. I modelli univariati che abbiamo preso in considerazione per confronto sono un MA(1,12) e un Holt-Winters con stagionalità additiva. Tra i due tipi di previsioni risulta sempre preferibile il metodo Holt-Winters. La qualità delle previsioni non peggiora passando da un lead di un mese a uno di due o tre mesi.

Dalle previsioni mensili è anche possibile ottenere previsioni trimestrali per aggregazione. Naturalmente, nel caso dei dati di elettricità è necessario integrare le stime a 1 e 2 lead con quelle a 3 lead di uno dei due metodi univariati considerati. Le previsioni trimestrali mostrano, rispetto a quelle mensili, una riduzione dell'errore di entità rilevante, dovuta però in parte al fatto che la serie trimestrale presenta una variabilità notevolmente inferiore a quella mensile. Solo con le stime trimestrali basate sui dati di elettricità si osserva anche una maggiore riduzione della varianza rispetto alle previsioni mensili; l'errore medio assoluto che si sarebbe commesso nella previsione del trimestre sarebbe stato di poco superiore ad un punto percentuale.

Abbiamo poi cercato di combinare le previsioni univariate con quelle basate sull'elettricità. Poiché il metodo ARIMA e quello Holt-Winters sono naturalmente molto correlati tra loro, abbiamo deciso di utilizzare solamente il secondo. Le combinazioni consistono in medie ponderate con pesi che sommano all'unità, dato che tutte le previsioni sono risultate non distorte. I vantaggi delle combinazioni si sono rivelati significativi sia per le previsioni ad 1 periodo in avanti, sia per quelle a 2 periodi. Tutte le formule impiegate per la combinazione, che non tengono conto della covarianza tra gli errori dei diversi metodi di previsione, hanno sempre fornito risultati più soddisfacenti del migliore singolo metodo previsivo. Inoltre, non esiste una chiara gerarchia qualitativa tra le formule di combinazione, con l'eccezione piuttosto disarmante della semplice media aritmetica che fornisce sempre i migliori risultati.

Se si considera un orizzonte previsivo trimestrale, con i dati ottenuti per aggregazione dalle stime già combinate, è confermato un netto miglioramento nell'accostamento ai dati effettivi rispetto ai migliori risultati ottenuti con un singolo metodo; la radice dell'errore quadratico medio si riduce del 22 per cento e l'errore medio assoluto del 39.

Un altro interessante risultato ottenuto con l'impiego delle combinazioni è stato quello di riuscire a restringere significativamente l'ampiezza degli intervalli di confidenza. Rispetto alle stime basate sui dati di elettricità mensili, a parità del livello di significatività scelto, si ha una riduzione dell'ampiezza media degli intervalli di confidenza, anche se vi è un lieve aumento nel numero di casi che cadono al di fuori di questi; rispetto al numero di osservazioni la percentuale osservata è comunque sempre largamente inferiore al livello scelto di significatività. Per quanto riguarda le combinazioni a 2 mesi, rispetto alle stime basate sui dati quattordicinali di energia elettrica, si ottiene contemporaneamente sia una riduzione nell'ampiezza media degli intervalli, sia una diminuzione nel numero di osservazioni che cadono al di fuori di questi.

In questo studio si sono per la prima volta utilizzati i dati giornalieri di energia elettrica. Ci sembra che in complesso i risultati ottenuti siano soddisfacenti e costituiscano una buona base per stime mensili e trimestrali tempestive dell'indice della produzione industriale.

NOTE

(*) Sebbene la ricerca sia frutto di un lavoro congiunto l'attribuzione delle varie parti è la seguente: l'introduzione, le conclusioni e il paragrafo 2.2 sono stati scritti da G. Bodo; l'appendice B e il paragrafo 2.1 da L. F. Signorini; i paragrafi 1 e 3 e l'appendice A da A. Cividini. Ringraziamo C.A. Bollino per utili commenti e M. Olivieri per il prezioso aiuto fornitoci nella preparazione del testo e nelle elaborazioni statistiche.

(1) Per comodità, in tutta l'esposizione con il termine "industria" si fa riferimento all'industria in senso stretto, al netto cioè dell'edilizia.

(2) Il modello è il seguente:

$$/2.1/ \quad \bar{X}_t = A(X_t - F_{t-s}) + (1 - A)(\bar{X}_{t-1} + T_{t-1})$$

$$/2.2/ \quad F_t = D(X_t - \bar{X}_t) + (1 - D) F_{t-s}$$

$$/2.3/ \quad T_t = C(\bar{X}_t - \bar{X}_{t-1}) + (1 - C) T_{t-1}$$

dove $0 < A < 1$, $0 < C < 1$, $0 < D < 1$

e dove s rappresenta il periodo della stagionalità; nel nostro caso, ovviamente, $s=12$. Il termine \bar{X} indica il livello della serie, T il trend e F la stagionalità. L'inizializzazione è stata compiuta ponendo, per $j = 1, \dots, 12$:

$$/2.4/ \quad F_j = X_j - 1/s \sum_{k=1}^s X_k$$

$$/2.5/ \quad \bar{X}_j = 1/s \sum_{k=1}^s X_k$$

$$/2.6/ \quad T_j = 0$$

- (3) Nella stima con l'utilizzo della griglia abbiamo calcolato la funzione di perdita quadratica per tutto l'insieme dei valori dei tre parametri A, C e D variandone i valori di .1. Una volta scelta la terna dei valori che minimizza la funzione di perdita abbiamo proceduto ad una nuova ricerca nell'interno a passi di .01. Abbiamo poi seguito lo stesso metodo fino ad incrementi di .001 nel valore dei parametri. Naturalmente l'applicazione di questa procedura non assicura necessariamente il raggiungimento di un punto di minimo globale.
- (4) Per tutte le previsioni a 1 periodo di distanza sono anche stati calcolati i test di Durbin-Watson e di Von Neumann-Hart per verificare l'ipotesi di assenza di autocorrelazione negli errori di previsione; in nessun caso l'ipotesi è stata respinta.
- (5) A tutti gli errori di previsione è stato applicato il test di Kolmogorov senza mai rifiutare l'ipotesi nulla di normalità, con livelli di significatività anche superiori al 10 per cento.
- (6) Vedi ad esempio Cipolletta-Marzano-Tresoldi-Signorini, (1985), Bodo-Signorini (1985b), Gutierrez-Saraceno (1987).
- (7) Il modello /3/ è sostanzialmente uguale a quello presentato in Bodo-Signorini (1985b); l'unica differenza di rilievo consiste nella rettifica dell'elettricità con i coefficienti suoi propri anziché con quelli della produzione.
- (8) Può essere opportuno esporre in maniera schematica le previsioni a cui si arriva per un qualsiasi mese. All'inizio della prima settimana del mese t sono disponibili le stime basate sui dati mensili di elettricità per il mese t-1, le previsioni dei metodi univariati per i mesi t-1, t e t+1 e le previsioni basate sulle inchieste congiunturali ISCO per il mese t-2 (vedi Bodo-Signorini, 1985b). Si ricorda inoltre che nella prima settimana del mese t l'indice di produzione noto è quello che si riferisce a t-3; solamente alla metà circa del mese si ha un dato provvisorio di produzione industriale per il mese t-2. Quasi contemporaneamente si rendono disponibili i dati sull'elettricità dei primi quattordici giorni che forniscono una valutazione per il mese t. Infine, al termine del mese è disponibile il dato definitivo della produzione industriale per t-2. La disponibilità dei dati è riassunta nella seguente tabella:

Settimana Previsioni con Inchieste Metodi Realizzazioni
del mese t elettricità congiunturali univariati

1 ^a (inizio)	Enel ^M _{t-1}	Isco _{t-2}	U _{t-1} , U _t , U _{t+1}	-
2 ^a (fine)	-	-	-	IP ^P ₋
3 ^a (inizio)	Enel ¹⁴ _t	-	-	-
4 ^a (fine)	-	-	-	IP ^D _{t-2}

- (9) Vedi Clemen (1986), Diebold-Pauly (1987), Granger-Ramanthan (1984), Mills-Stephenson (1985), Holden-Peel (1986).
- (10) I valori dei pesi non calcolabili nei primi periodi mediante le formule sono stati posti arbitrariamente uguali a 0.5.
- (11) Il coefficiente di correlazione osservato tra gli errori di previsione delle stime basate sui dati di elettricità e quelli del metodo Holt-Winters è molto basso (0.16) e non significativo ad 1 periodo di distanza; è, invece, più elevato (0.67) e significativo per 2 periodi di distanza. Anche in quest'ultimo caso, tuttavia, poichè teoricamente i due metodi sono indipendenti, abbiamo preferito ignorare la correlazione.

APPENDICE A

Calcolo degli intervalli di confidenza

Per il calcolo degli intervalli di confidenza si sono utilizzate le formule standard nel caso delle stime OLS del modello mensile e quattordicinale ENEL, e nel caso dei modelli ARIMA (ad esempio si veda Piccolo-Vitale (1984)). Per quanto riguarda il metodo Holt-Winters, invece, è stato necessario effettuare una serie di elaborazioni per valutare la varianza dell'errore di previsione a lead l , con $l \geq 1$.

Utilizzando il modello Holt-Winters additivo descritto dalle relazioni 2.1-2.3 della nota (2), come mostrato da Sweet (1985), si può ricavare l'errore di previsione a lead l mediante le seguenti relazioni:

$$\text{/A1/} \quad e_t(l) = \left[1 - \frac{B^l K(B,l)}{K(B)} \right] u_{t+l} \quad 1 \leq l < s$$

$$\text{/A2/} \quad K(B) = 1 - (1-A)B + AC \sum_{i=1}^s B^i + D[(1-A)-1]B^s + (1-A)(1-D)B^{s+1}$$

$$\text{/A3/} \quad K(B,l) = A[1 - (1-C)B + C^l(1-B)] \sum_{i=0}^{s-1} B^i + D(1-A)(1-B)B^{s-1}$$

dove:

u è il termine di errore ipotizzato bianco e gaussiano

B è l'operatore di ritardo unitario

s è il periodo della stagionalità.

La varianza dell'errore di previsione a lead l si ottiene, a partire dalle A1-A3 mediante le seguenti relazioni:

$$/A4/ \quad \alpha(l, B) = 1 - \frac{B^l K(B, l)}{K(B)} = \sum_{k=0}^{\infty} \alpha(l)_k B^k$$

$$/A5/ \quad V(l) = \sigma^2 \sum_{k=0}^{\infty} \alpha(l)_k^2 \quad 1 \leq l < s$$

dove σ^2 è la varianza del termine di errore u.

In particolare, per $l=1$ la /A4/ si semplifica nella

$$/A6/ \quad \alpha(B) = \frac{(1-B)(1-B^s)}{K(B)}$$

Le formule riportate sono valide solo se converge la serie

$$\sum_{k=0}^{\infty} \alpha(l)_k^2$$

cioè se il modello Holt-Winters è stabile, in caso contrario la varianza dell'errore di previsione non è finita. Per una trattazione completa delle condizioni di stabilità si veda ad esempio, Mc Clain e Thomas (1973).

APPENDICE B

La rettifica nella previsione della produzione industriale

Una grandezza-flusso è definita come una quantità (prodotta, scambiata, consumata, etc.) nell'unità di tempo; è naturale che sorgano inconvenienti se l'unità di tempo utilizzata non è fissa. L'unità di tempo a cui si riferisce l'indice della produzione industriale, il mese, contiene a seconda dei casi da 28 a 31 giorni di calendario e da 19 a 23 giorni lavorativi (dal lunedì al venerdì, festività escluse). Per i giorni lavorativi il campo di variazione è quindi di oltre il 20 per cento. Sebbene per motivi istituzionali o contabili si faccia generalmente riferimento ai mesi di calendario, è evidente che il fenomeno economico sottostante - cioè l'intensità del flusso - viene misurato più accuratamente se si fa ricorso a una definizione più rigorosa dell'unità di tempo. Per questo motivo si utilizza talvolta la produzione media giornaliera, ossia la produzione divisa per il numero dei giorni del mese. Questa operazione viene generalmente chiamata "rettifica".

In particolare è di solito opportuno usare dati rettificati in esercizi di previsione. Qualunque modello della produzione industriale si basa, almeno idealmente, su un qualche fenomeno reale del quale le relazioni statistiche stimate sono l'astrazione sommaria. E' plausibile che i fenomeni in questione abbiano a che fare con l'intensità del flusso rigorosamente definita, e prescindano da dati del tutto convenzionali e artificiali quali i confini tra un mese e l'altro. La previsione dovrebbe quindi risultare in generale più accurata usando la produzione media giornaliera. Qualora interessino previsioni di dati grezzi, è sempre possibile ottenerle compiendo l'operazione inversa alla rettifica sulle previsioni di dati rettificati.

Queste considerazioni si applicano in primo luogo ai modelli univariati. Se è l'intensità, cioè il ritmo, del flusso produttivo ad essere generata da un processo stocastico - anziché il suo integrale su un periodo di lunghezza variabile -, allora

l'uso di dati grezzi comporta errori di identificazione e peggiora la qualità delle stime. Gli errori di identificazione nascono dalla comparsa di autocorrelazioni spurie dovute all'oscillazione più o meno regolare della lunghezza dei mesi e all'interazione con quella delle settimane. Se si stima un modello sui dati grezzi si trattano in modo impreciso e stocastico (termini AR o MA a ritardi "spuri") fenomeni deterministici e perfettamente conosciuti (calendario): le previsioni sono quindi inefficienti. Per dettagli e verifiche vedi Bodo-Signorini, 1985b.

Nella stima della produzione industriale sulla base dei dati elettrici il problema della rettifica scompare se elettricità e produzione sono legate da una semplice relazione del tipo $Y_t = bE_t + \text{errore casuale}$. In questo caso usare dati grezzi o rettificati è indifferente (a prescindere dall'eteroschedasticità), perchè si può passare dagli uni agli altri semplicemente moltiplicando per lo stesso coefficiente sia il primo sia il secondo membro della relazione. Tuttavia un modello così semplice non funziona per motivi ben noti e ampiamente discussi nella letteratura sull'argomento: è necessario introdurre differenziazioni, costanti, dummies stagionali, termini ritardati etc.. Se la relazione "vera" si riferisce a flussi rigorosamente definiti, cioè a dati rettificati, l'uso dei dati grezzi comporta inconvenienti analoghi a quelli relativi ai modelli univariati. Anche in questo caso Bodo e Signorini verificano empiricamente la superiorità dei modelli basati su dati rettificati.

Una complicazione è data dal fatto che i coefficienti di rettifica delle due serie non sono gli stessi. Per l'indice della produzione la rettifica è legata naturalmente solo all'attività industriale: sabati, domeniche e giorni festivi "pesano" secondo l'importanza delle industrie a ciclo continuo, cioè attualmente circa il 15% di un giorno lavorativo ordinario. Per l'elettricità i giorni non lavorativi pesano di più per due motivi: primo, a causa dei consumi extraindustriali (domestici, etc.), che presumibilmente sono distribuiti nell'arco della settimana in modo più uniforme rispetto all'attività produttiva; secondo,

perchè le lavorazioni a ciclo continuo, di solito caratterizzate da un'alta intensità elettrica, incidono molto di più sui consumi elettrici che sulla produzione (45% contro 15% circa).

Nel caso della produzione industriale il peso di ciascun giorno della settimana ai fini della rettifica deve essere attribuito a priori. Invece per l'elettricità i pesi in questione possono essere valutati empiricamente usando i dati giornalieri sull'energia immessa in rete. Più specificamente, si può calcolare per ciascun giorno della settimana la media - su tutte le settimane disponibili - del rapporto tra l'immissione di elettricità in quel giorno e quella del mercoledì (scelto come paradigma di giorno "pieno" perchè generalmente è il giorno in cui l'immissione elettrica tocca il massimo). Le stime ottenute con questo metodo confermano che il peso dei giorni festivi è di gran lunga maggiore del 15% postulato per la produzione: il valore si pone attorno al 70% per la domenica e arriva all'85% per il sabato, presumibilmente a causa delle attività lavorative (soprattutto terziarie) che non si interrompono in tale giorno.

La conseguenza teorica di questa diversità è che il modello elettricità-produzione deve contenere un certo numero di correttivi. Per precisarli occorre esaminare in maggior dettaglio i fondamenti della relazione.

In qualunque istante il flusso di energia elettrica, consta di tre componenti E_1 , E_2 , E_3 , corrispondenti rispettivamente ai consumi delle industrie a ciclo continuo, ai consumi delle industrie a ciclo discontinuo e ai consumi extraindustriali (si trascurano qui le perdite). Supponiamo che valgano le seguenti relazioni tecniche (esatte):

$$/B1/ \quad E_1 = a_1 + b_1 Y_1;$$

$$/B2/ \quad E_2 = a_2 + b_2 Y_2;$$

$$/B3/ \quad E_3 = a_3 + b_3 T;$$

dove Y_1, Y_2 , rappresentano la produzione rispettivamente delle industrie a ciclo continuo e discontinuo, mentre T è la variabile (o il vettore di variabili) che determina i consumi extraindustriali E_3 (temperatura, etc.).

Supponiamo inoltre (per ora) che nel corso del mese t le attività Y_1, Y_2 , siano costanti, eccetto che Y_2 si ferma, per definizione, nei giorni non lavorativi. Analogamente, supponiamo che la temperatura e le altre variabili non varino, cosicchè eventuali cambiamenti inframensili dei consumi extraindustriali E_3 siano unicamente dovute alla presenza di una "stagionalità settimanale" fissa loro propria. Siano l_{1t} il numero giorni del calendario del mese; l_{2t} il numero dei giorni lavorativi; l_{3t} il numero dei giorni ponderati ai fini dei consumi extraindustriali ("mercoledì-equivalenti"). Il consumo totale di energia nel mese, misurato in unità opportune, sarà:

$$\begin{aligned} /B4/ \quad E^M_t &= l_{1t}E_{1t} + l_{2t}E_{2t} + l_{3t}E_{3t} = \\ &= l_{1t} a_1 + l_{2t} a_2 + l_{3t} a_3 + (l_{1t} b_{1t} p + l_{2t} b_{2t} (1-p)) Y + l_{3t} b_{3t} T \end{aligned}$$

dove $Y = Y_1 + Y_2$ e $0 < p = Y_1/Y < 1$ (d'ora in avanti il suffisso temporale t sarà sottinteso). Si noti che qui Y è la produzione totale di un giorno non festivo del mese t , ossia la produzione media giornaliera, o produzione rettificata che dir si voglia. Il coefficiente di rettifica della produzione industriale l_y , definito come il rapporto tra la produzione mensile grezza $Y_M = l_1 Y_1 + l_2 Y_2$ e la produzione media giornaliera Y , è dato ovviamente da

$$/B5/ \quad l_y = (Y_M/Y) = l_1 p + l_2 (1-p)$$

Date le ipotesi /B1/ - /B3/, si può esprimere il flusso

di elettricità di un giorno "pieno" (mercoledì non festivo) come:

$$/B6/ \quad E = a + bY + E_3$$

dove $a = a_1 + a_2$ e $b = pb_1 + (1-p)b_2$. Definendo inoltre q come quota pertinente alle industrie a ciclo continuo dei consumi elettrici industriali, sempre in un giorno "pieno":

$$/B7/ \quad q = E_1 / (E_1 + E_2) = (a_1 + b_1 Y_1) / (a + bY)$$

si ha dopo alcuni passaggi:

$$/B8/ \quad E^M = l_e a + b l_e Y + l_3 a_3 + b_3 l_3 T$$

dove $l_e = l_1 q + l_2 (1-q)$ (coefficiente di rettifica ponderato dell'elettricità industriale). Invertendo la /B8/ per esprimere Y in funzione di E si ottiene infine:

$$/B9/ \quad Y \equiv Y^M / l_Y = A + B(E^M / l_e) + C(l_3 / l_e) T + D(l_3 / l_e)$$

$$\text{dove: } A = b^{-1} a$$

$$B = b^{-1}$$

$$C = -b^{-1} b_3$$

$$D = -b^{-1} a_3$$

Realisticamente, la relazione tra elettricità e produzione deve essere considerata stocastica; alla /B9/ andrà perciò aggiunto un errore casuale sul quale si faranno le consuete ipotesi (media nulla, assenza di autocorrelazione, omoschedasticità e ortogonalità rispetto alle variabili indipendenti). Inoltre sarà necessario, come si è detto, introdurre nella relazione la produzione (rettificata) ritardata, nonché dummies stagionali.

Come si vede dalla /B9/, la presenza di costanti e con-

sumi extraindustriali richiederebbe, oltre alla rettifica dell'elettricità immessa, ulteriori correzioni, date dal rapporto tra i coefficienti di rettifica dell'elettricità extraindustriale e industriale; tale rapporto dovrebbe essere introdotto sia come variabile indipendente a sé stante, sia per la "rettifica" della variabile (o delle variabili) T. L'effettiva rilevanza di queste correzioni è comunque una questione empirica. Le nostre stime miglioravano solo marginalmente; perciò si è scelto di semplificare il modello esposto nel testo escludendo tali correzioni. Non è escluso tuttavia che esse possano rivelarsi più rilevanti in futuro.

Ovviamente né p né q sono costanti (anzi, data la presenza di consumi "fissi", se p è costante non può esserlo q e viceversa). Nascono quindi problemi nel calcolo di l_y e l_e . Sarebbe necessario rettificare separatamente la componente continua e quella discontinua sia di Y che di E ; ma i relativi dati si conoscono solo con un certo ritardo per quanto riguarda Y , e con un forte ritardo e comunque mai a livello mensile (solo a livello semestrale) per E . E' necessario quindi introdurre delle approssimazioni. Per la produzione industriale l'ultimo valore dell'indice si può rettificare in via provvisoria calcolando l_y sulla base del valore di p nell'anno-base; non appena i valori degli indici disaggregati si rendono disponibili, si può eseguire la rettifica corretta. L'esperienza (Bodo-Signorini, 1985a) dimostra che lo scostamento della rettifica provvisoria da quella definitiva è minimo, quasi impercettibile. Anche per q (elettricità) si possono usare i valori relativi a un certo anno di riferimento, ed è naturale utilizzare ancora una volta il 1980 per uniformità con la produzione; tuttavia in questo caso non è possibile calcolare la rettifica nel modo teoricamente corretto per la mancanza di dati di cui si è detto. Ci si deve quindi accontentare di una rettifica approssimativa, confidando - sulla base dei risultati ottenuti con la produzione - che le relative distorsioni non siano eccessive.

BIBLIOGRAFIA

- BODO, G. (1984) - "Un modello previsivo dell'andamento della produzione industriale", Rapporto Progetto DESEC n° 12.
- BODO, G. - SIGNORINI, L.F. (1985a), "Ricostruzione storica e depurazione stagionale degli indici di produzione industriale", Banca d'Italia, Bollettino Statistico, n° 3-4.
- BODO, G. - SIGNORINI, L. F. (1985b), "Uno schema per la previsione a breve termine della produzione industriale". Temi di discussione, n° 55.
- BODO, G. - SIGNORINI, L. F. (1987), "Short-term forecasting of the industrial production index", International Journal of Forecasting, n. 3.
- CIPOLLETTA, I., MARZANO, G., TRESOLDI, C., SIGNORINI, L. F. (1985), "L'analisi congiunturale attraverso i dati elettrici", in: Enel, Rapporto sull'attività del Comitato Tecnico-Scientifico per la Programmazione Elettrica nel biennio 1983-84, Roma.
- CLEMEN, R.T. (1986), "Linear constraints and the efficiency of combined forecasts", Journal of Forecasting, n° 1.
- DIEBOLD, F.X. - PAULY P. (1987), "Structural change and the combination of forecasts", Journal of Forecasting, n° 1.
- GRANGER, C.W.J. - NEWBOLD, P. (1977), Forecasting economic time series, New York, Academic Press.
- GRANGER, C.W.J - RAMANANTHAN, R. (1984), "Improved methods of combining forecasts", Journal of Forecasting, n° 3.
- GUTIERREZ, L. - SARACENO, P. (1987 "Un indicatore di attività in tempo reale per orientare le decisioni di politica economica", Contributi di Ricerca IRS, n° 4.
- HOLDEN, K. - PEEL, D.A. (1986), "An empirical investigation of combination of economic forecasts", Journal of Forecasting, n° 4.
- MC CLAIN, J.O. - THOMAS, L.J. (1973), "Response-variance trade-off in adaptive forecasting", Operations Research, 21.
- MAKRIDAKIS, S. - et. al. (1982), "The accuracy of extrapolation methods: results of a forecasting competition", Journal of Forecasting, n. 1.
- MAKRIDAKIS, S. - WINKLER, R. L. (1983), "The combination of forecasts", Journal of the Royal Statistical Society, Series A.
- MILLS, T.C. - STEPHENSON M.J. (1985), "Forecasting contemporaneous aggregates and the combination of forecasts: the case

of the U.K. monetary aggregates", Journal of Forecasting, n° 4.

PICCOLO, D. - VITALE, C. (1984), Metodi statistici per l'analisi economica, Bologna, Il Mulino.

SWEET, A.L. (1985), "Computing the variance of the forecast error for the Holt-Winters seasonal model", Journal of Forecasting, n° 2.

ELENCO DEI PIÙ RECENTI TEMI DI DISCUSSIONE (*)

- n. 91 — *La disoccupazione in Italia: un'analisi con il modello econometrico della Banca d'Italia*, di G. BODO - I. VISCO (luglio 1987).
- n. 92 — *L'Italia e il sistema monetario internazionale dagli anni '60 agli anni '90 del secolo scorso*, di M. ROCCAS (agosto 1987).
- n. 93 — *Reddito e disoccupazione negli Stati Uniti e in Europa: 1979-1985*, di J. C. MARTINEZ OLIVA (agosto 1987).
- n. 94 — *La tassazione e i mercati finanziari*, di G. ANCIDONI - B. BIANCHI - V. CERIANI - P. CORRAGGIO - A. DI MAJO - R. MARCELLI - N. PIETRAFESA (agosto 1987).
- n. 95 — *Una applicazione del filtro di Kalman per la previsione dei depositi bancari*, di A. CIVIDINI - C. COTTARELLI (ottobre 1987).
- n. 96 — *Macroeconomic Policy Coordination of Interdependent Economies: the Game-Theory Approach in a Static Framework*, di J. C. MARTINEZ OLIVA (ottobre 1987).
- n. 97 — *Occupazione e disoccupazione: tendenze di fondo e variazioni di breve periodo*, di P. SYLOS LABINI (novembre 1987).
- n. 98 — *Capital controls and bank regulation*, di G. GENNOTTE - D. PYLE (dicembre 1987).
- n. 99 — *Funzioni di costo e obiettivi di efficienza nella produzione bancaria*, di G. LANCIOTTI - T. RAGANELLI (febbraio 1988).
- n. 100 — *L'imputazione di informazioni mancanti: una sperimentazione*, di L. CANNARI (marzo 1988).
- n. 101 — *Esiste una curva di Beveridge per l'Italia? Analisi empiriche della relazione tra disoccupazione e posti di lavoro vacanti (1980-1985)*, di P. SESTITO (marzo 1988).
- n. 102 — *Alcune considerazioni sugli effetti di capitalizzazione determinati dalla tassazione dei titoli di Stato*, di D. FRANCO - N. SARTOR (luglio 1988).
- n. 103 — *La coesione dello SME e il ruolo dei fattori esterni: un'analisi in termini di commercio estero*, di L. BINI SMAGHI - S. VONA (luglio 1988).

(*) I «Temi» possono essere richiesti alla Biblioteca del Servizio Studi della Banca d'Italia.

