

BANCA D'ITALIA

Temi di discussione

del Servizio Studi

**Una applicazione del filtro di Kalman
per la previsione dei depositi bancari**

di Andrea Cividini e Carlo Cottarelli



Numero 95 - Ottobre 1987

BANCA D'ITALIA

Temi di discussione

del Servizio Studi

**Una applicazione del filtro di Kalman
per la previsione dei depositi bancari**

di Andrea Cividini e Carlo Cottarelli

Numero 95 - Ottobre 1987

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 disponibilità di informazioni accurate e tempestive sull'andamento dei principali aggregati monetari e creditizi costituisce un presupposto essenziale per una efficiente conduzione della politica monetaria.

Gli autori espongono una applicazione del filtro di Kalman per la depurazione di informazioni provvisorie sulla crescita dei depositi dalla componente di errore.

I risultati ottenuti sono soddisfacenti e mostrano come sia possibile ridurre gli errori dei dati provvisori di circa il 15-20 per cento e, in misura anche superiore, la loro dispersione.

INDICE

1 - Introduzione	pag.	5
2 - Il campione decadale della Banca d'Italia	pag.	7
3 - Metodi di stima preliminari	pag.	12
4 - Il filtro di Kalman	pag.	19
5 - La procedura utilizzata per l'applicazione del filtro di Kalman	pag.	23
6 - Risultati della applicazione del filtro di Kalman	pag.	24
7 - Conclusioni	pag.	46
Note	pag.	48
Bibliografia	pag.	50

1 - Introduzione (*)

Nei principali paesi industrializzati le banche centrali fissano periodicamente, con diverso grado di dettaglio e orizzonte temporale, dei piani finanziari, definiti talvolta "obiettivi intermedi" della politica monetaria, utilizzati nel corso del periodo come punto di riferimento per l'azione svolta quotidianamente sui mercati finanziari e valutari. Dallo scostamento delle variabili osservate, per esempio la quantità di moneta o di credito, dal sentiero programmato possono derivare importanti informazioni sul comportamento reale e finanziario degli operatori; queste nuove informazioni possono indurre a rettificare la manovra degli strumenti a disposizione delle autorità monetarie per il perseguimento degli obiettivi finali (1). La disponibilità di informazioni accurate e tempestive sull'andamento dei principali aggregati monetari e creditizi costituisce quindi un presupposto essenziale per una efficiente conduzione della politica monetaria. Una cattiva qualità delle informazioni utilizzate presenta costi elevati perché può condurre all'uso inappropriato degli strumenti e quindi a deviazioni degli obiettivi finali (il reddito reale e nominale, l'equilibrio esterno, ecc.) dai valori prefissati. Questi costi vanno contrapposti a quelli affrontati nella raccolta delle informazioni che, anche se ridotti dall'innovazione tecnologica, restano pur sempre elevati.

Dalla necessità di conciliare l'esigenza di informazioni tempestive sull'andamento degli aggregati monetari con quella di

(*) Una precedente versione di questo lavoro è stata presentata al VII Convegno COMPSTAT, tenuto a Roma, presso l'Università "La Sapienza", dall'1 al 5 settembre 1986. Si ringraziano il dr. Ignazio Visco per i suggerimenti forniti nel corso dell'intero lavoro e il dr. Giorgio Bodo per gli interessanti commenti.

contenere il costo dell'informazione che grava sul sistema è derivato l'uso da parte delle banche centrali, e delle agenzie governative in generale, di segnalazioni provvisorie che forniscano prime indicazioni sugli sviluppi congiunturali in atto. Queste segnalazioni contengono di solito un disturbo elevato, sia perché spesso non si riferiscono all'intero sistema, sia perché la tempestività comporta quasi inevitabilmente errori e incompletezze nella registrazione e nella trasmissione dei dati di base.

L'analisi statistica fornisce diversi strumenti per depurare le informazioni provvisorie dalla componente di errore, consentendo di estrarre il segnale da utilizzare da parte del policy-maker. Tra i più interessanti sviluppi manifestatisi in questo settore, occorre segnalare la recente applicazione all'analisi delle serie storiche economiche del filtro di Kalman (cfr., per esempio, Conrad e Corrado (1979), Engle e Watson (1985), Cuthbertson e Taylor (1985), Bordignon e Trivellato (1986)). Il filtro di Kalman consente di combinare ottimamente le informazioni sull'innovazione di un processo, incorporate nelle segnalazioni provvisorie, con quelle derivabili dall'analisi delle caratteristiche dinamiche del processo stesso (attraverso un modello ARIMA o un modello strutturale (cfr. Harvey (1981))). Questo lavoro presenta una applicazione del filtro di Kalman alla stima dell'andamento dei depositi bancari delle aziende di credito italiane. Questo andamento è di importanza essenziale per la politica monetaria dato che in Italia i depositi bancari rappresentano l'80 per cento circa dello stock di moneta del settore privato (nella definizione M2), una delle variabili intermedie utilizzate dalle autorità monetarie (Ciampi (1987)).

In relazione a questo problema, il paragrafo 2 illustra le caratteristiche delle segnalazioni provvisorie sulla crescita dei depositi attualmente raccolte dalla Banca d'Italia. Da queste informazioni, disponibili decadalmente con un ritardo di cinque giorni lavorativi rispetto alla data di riferimento, ven-

gono derivate, attraverso metodi estrapolativi, prime stime sulla crescita della raccolta bancaria nel corso di ogni mese (par. 3). Come descritto nel paragrafo 4 queste stime possono essere efficientemente combinate, attraverso il filtro di Kalman, con le previsioni di modelli che descrivono le caratteristiche dinamiche delle serie considerate. Nel paragrafo 5 si descrive la strategia seguita per l'applicazione empirica e nel paragrafo 6 si riportano i risultati: nell'esercizio particolare attenzione è stata prestata all'utilizzo per la previsione unicamente delle informazioni disponibili in ogni periodo considerato; di conseguenza le stime dei parametri del modello sono state effettuate utilizzando i dati disponibili fino al dicembre 1984, mentre il biennio 1985-86 è stato considerato per valutare la capacità previsiva del filtro di Kalman. Per valutare la sensibilità del filtro di Kalman a variazioni nella "qualità" delle segnalazioni provvisorie, l'analisi empirica viene svolta per due gruppi di banche: da un lato le aziende di credito maggiori, grandi e medie (che hanno una rappresentatività nel campione decadale che oscilla intorno al 95 per cento), dall'altro le aziende piccole e minori (rappresentate per circa il 50 per cento) (2). I risultati indicano la possibilità di ridurre l'errore di osservazione presente nei dati provvisori di un importo non trascurabile per l'analisi congiunturale del mercato monetario. Infine, il lavoro viene completato con la presentazione di un metodo, basato sempre sul filtro di Kalman, che consente di prevedere l'andamento nel corso del mese dei depositi bancari avendo a disposizione informazioni provvisorie inerenti unicamente all'evoluzione dell'aggregato nelle prime due decadi.

2 - Il campione decadale della Banca d'Italia

Fino al 1978 la Banca d'Italia riceveva informazioni sulla consistenza alla fine di ogni mese dei depositi bancari unicamente attraverso le segnalazioni statistiche richieste a fini di Vigilanza sulla base della legge bancaria (3). Queste informa-

zioni sono estremamente dettagliate e, tenendo conto dei ritardi nella trasmissione dei dati e nei tempi di elaborazione e di controllo, risultano disponibili con un ritardo compreso tra i 40 e i 60 giorni dalla data di riferimento. Per ridurre il ritardo nella disponibilità delle informazioni sul comportamento delle banche, nel settembre del 1978 vennero introdotte le segnalazioni decadali. Queste segnalazioni sono relative, oltre che ad alcuni tassi di interesse, alle principali voci di bilancio delle banche (essenzialmente i depositi, i prestiti in lire e valuta, i titoli). Le prime informazioni decadali (che nel resto del lavoro vengono chiamate "primi invii") sono trasmesse con un ritardo di cinque giorni lavorativi dalla data di riferimento con periodicità decadale (4). Con un ritardo di circa 15 giorni lavorativi le aziende comunicano eventuali rettifiche ai primi invii decadali; nel presente lavoro i dati rettificati sono chiamati "secondi invii". I dati comunicati a fini di Vigilanza sono chiamati "dati definitivi".

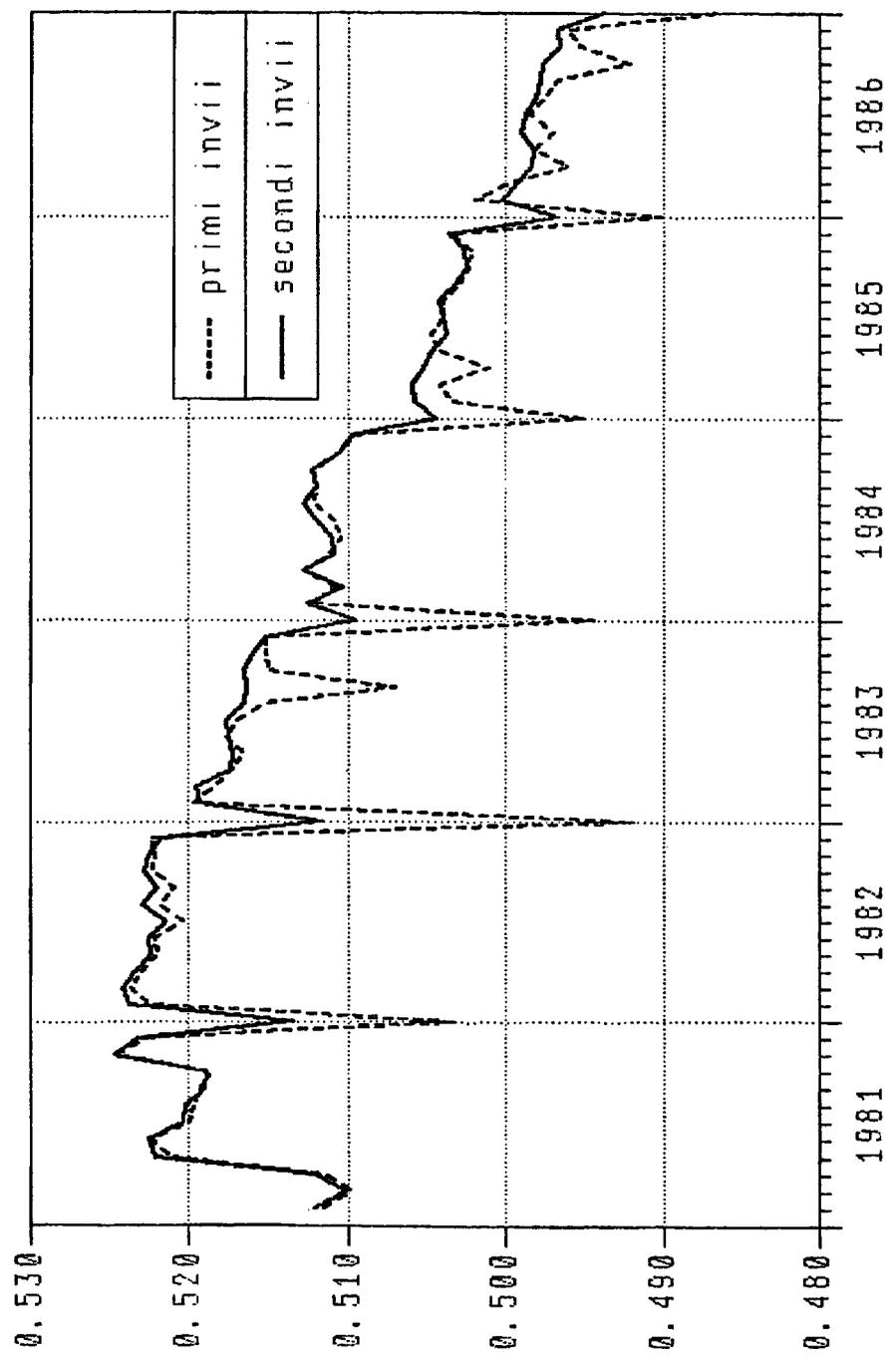
La qualità delle statistiche decadali è cresciuta nel tempo per tre motivi: l'ampliamento del campione che invia le segnalazioni, l'aumento della rappresentatività dei dati decadali di ogni azienda rispetto al dato definitivo (dovuta ad una progressiva maggiore copertura degli sportelli segnalati da ciascuna azienda) e un naturale miglioramento, dovuto agli avanzamenti tecnologici nella rilevazione dei dati, che ha consentito una riduzione degli errori. Nonostante ciò, le segnalazioni decadali presentano ancor oggi un margine di errore non trascurabile che risulta naturalmente maggiore per i primi invii.

Occorre inoltre osservare che, anche attualmente, non tutte le aziende inviano le segnalazioni decadali: il campione decadale comprendeva al dicembre 1986 88 aziende rappresentanti l'80 per cento circa del totale dei depositi. Nel presente lavoro si utilizzano i dati relativi soltanto a 85 aziende, avendo escluso le aziende entrate nel campione solo nel periodo più recente. Le 85 aziende individuate non rappresentano tuttavia un campione

omogeneo: esso comprende infatti 31 aziende definite nella classificazione dimensionale della Banca d'Italia maggiori, grandi e medie (32 aziende in totale) e un istituto di categoria, con una rappresentatività sul totale dei depositi dei corrispondenti gruppi del 95 per cento circa; le restanti 53 aziende sono di dimensione piccola o minore e rappresentano il 50 per cento circa dei depositi corrispondenti. Dato che per la banca centrale assume interesse non solo l'andamento del totale, ma anche quello per gruppi, nel lavoro vengono presentate stime distintamente per i due gruppi delle aziende maggiori, grandi e medie e degli istituti di categoria (nel resto del lavoro chiamate sinteticamente grandi) e per le piccole e minori (chiamate sinteticamente piccole).

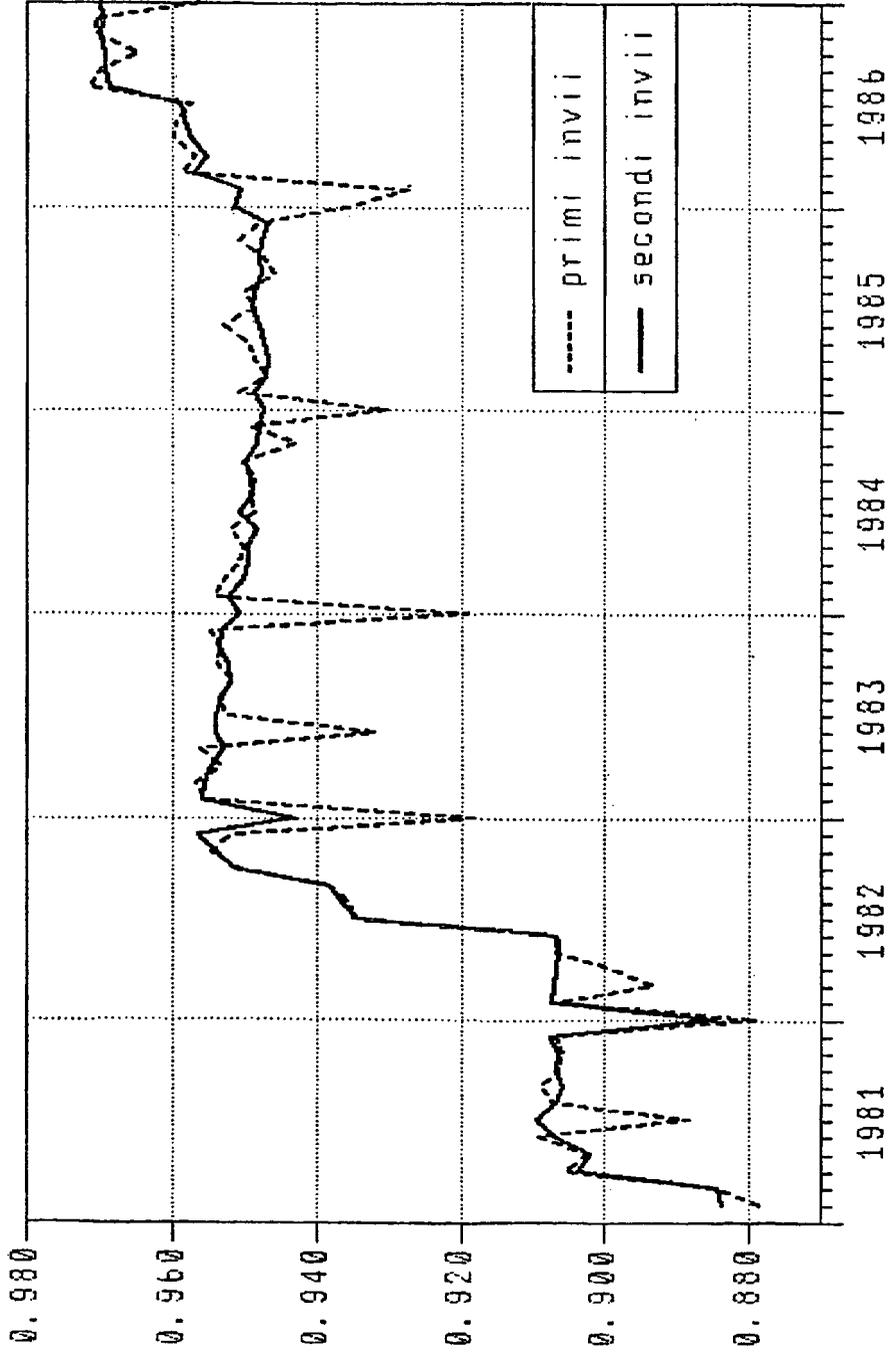
Come illustrato dalle figure 1 e 2, il rapporto tra dato decennale e dato definitivo (quello rilevato dalle statistiche di Vigilanza) si è andato riducendo nel tempo per le aziende piccole (fig. 1) come riflesso della progressiva perdita di mercato delle aziende incluse nel campione. Per le aziende grandi (fig. 2) invece la rappresentatività è tendenzialmente stabile dalla data in cui il campione raggiunge una massima dimensione (32 aziende). Le due figure evidenziano anche che la rappresentatività è molto più stabile se calcolata sui secondi invii rispetto ai primi, che risentono di un maggiore grado di incompletezza e di erraticità. E' infine chiaro che la rappresentatività dei dati decennali non è indipendente dal mese considerato; in dicembre la rappresentatività presenta un calo molto accentuato dovuto essenzialmente alle difficoltà connesse con la chiusura dei conti di fine anno. Si nota tuttavia che il calo nella rappresentatività è più attenuato negli ultimi anni, in particolar modo per i secondi invii, per il già citato miglioramento nella tecnologia di rilevazione dei dati.

RAPPRESENTATIVITA' DEL CAMPIONE DECADEALE
(aziende piccole) (1) Fig. 1



(1) Dall'ottobre 1981 il gruppo di aziende considerate costituisce un campione chiuso di 53 aziende di credito.

RAPPRESENTATIVITA' DEL CAMPIONE DECADEALE
(aziende grandi) (1) Fig. 2



(1) Dal settembre 1982 il gruppo di aziende considerate costituisce un campione chiuso di 32 aziende di credito.

3 - Metodi di stima preliminari

Il problema considerato in questo lavoro è quello di stimare la consistenza dei depositi alla fine dei mesi $t-1$ e t sulla base delle informazioni disponibili 5 giorni lavorativi dopo la fine del mese t . Anche se fattori casuali (ad esempio scioperi dei dipendenti bancari) rendono difficile una precisa definizione del set informativo, è presumibile che a tale data siano di norma disponibili le seguenti informazioni:

- lo stock definitivo dei depositi al tempo $t-2$ (R_{t-2});
- il secondo invio dei dati decadali relativo alla terza decade del mese $t-1$ ($D2_{t-1}$);
- il primo invio relativo alla terza decade del mese t ($D1_t$).

Date queste informazioni, sono considerati nel seguito di questo paragrafo due metodi di stima preliminare della consistenza dei depositi al tempo t e $t-1$. Questi metodi utilizzano unicamente le informazioni decadali relative alla terza decade di ogni mese; alla fine del paragrafo 6 si utilizzano invece le informazioni inframensili per prevedere lo stock dei depositi esistente a fine mese.

Il primo metodo è basato su un semplice riproporzionamento del flusso dei depositi segnalato nel mese dal campione decadale. Più precisamente il flusso dei depositi nel periodo $t-1$ (v_{t-1}^K) è calcolato come:

$$/3.1/ \quad v_{t-1}^K = (D2_{t-1} - D2_{t-2})/p_t$$

e la consistenza dei depositi alla fine del mese $t-1$ è ottenuta come:

$$/3.2/ K_{t-1} = R_{t-2} + V_{t-1}^K$$

dove K_{t-1} indica la stima della consistenza dei depositi effettuata al tempo t relativamente al tempo $t-1$ (stima rivista) e p_t è un coefficiente di riproporzionamento calcolato come segue:

$$/3.3/ p_t = (\bar{R}_{t-2} - \bar{R}_{t-14}) / (R_{t-2} - R_{t-14})$$

dove \bar{R} indica la consistenza dei depositi delle aziende incluse nel campione decadale. Il metodo descritto equivale a calcolare la rappresentatività delle aziende del campione sul flusso totale dei depositi per i dodici mesi precedenti l'ultimo dato definitivo e di riproporzionare il flusso rilevato dalle statistiche decadali in base a tale rappresentatività. Il dato di consistenza è calcolato quindi addizionando il flusso riproporzionato all'ultima consistenza dei depositi disponibile (5). Si noti che non tutti i dati decadali sono utilizzati per effettuare il calcolo descritto. Si è infatti deciso di escludere mese per mese dal campione decadale le aziende che presentassero variazioni nel mese di importo superiore al 15 per cento (del 35 per cento per i mesi di dicembre e gennaio): variazioni superiori a tale importo, sulla base della esperienza passata, sono di norma attribuibili a semplici errori nella trasmissione dei dati. Il campione considerato di mese in mese non è quindi costante, ma varia in base alle aziende escluse (6).

Il metodo descritto è in larga misura "ad hoc", in quanto non è derivato esplicitamente dalla formulazione di un modello che descriva la dinamica del campione rispetto a quella del sistema. Esso presenta tuttavia due importanti caratteristiche: innanzitutto, dato che il calcolo della consistenza dei depositi al tempo t è basato sulla stima di un flusso e non della consistenza direttamente, il metodo utilizza tutte le informazioni disponibili al tempo t , in particolare quella della consistenza dei depositi al tempo $t-2$. In secondo luogo, dato che la rappresen-

tatività è calcolata ugualmente su dati di flusso, il metodo consente di cogliere parzialmente la dinamica del processo che regola il mutamento della rappresentatività stessa. Dato che nel caso presentato tale rappresentatività costituisce la quota di mercato delle aziende incluse rispetto a quelle escluse, è probabile che questa sia legata a una dinamica ben precisa, connessa ai mutamenti in corso nel mercato dei depositi bancari. Il metodo di calcolo presentato costituisce quindi una via intermedia tra il considerare la rappresentatività su stock come costante e una esplicita modellazione della dinamica che regola le quote di mercato.

La stima del flusso dei depositi al tempo t (v_t^H) è ricavata come segue:

$$/3.4/ \quad v_t^H = (D1_t - D2_{t-1})/p_t$$

e la consistenza dei depositi è ottenuta come:

$$/3.5/ \quad H_t = K_{t-1} + v_t^H$$

dove H_t è la stima iniziale dei depositi al tempo t effettuata al tempo t stesso (7).

Il secondo metodo è basato sull'ipotesi che i dati del campione decadale siano legati a quelli del totale del sistema dalla seguente relazione:

$$/3.6/ \quad \nabla R_t^i = a + b \nabla D_t^i + u_t \quad i=1,2,3 \quad t=1,\dots,T$$

dove l'indice i identifica la decade cui il dato si riferisce, t indica il mese e ∇ indica la variazione dell'aggregato nel corso della decade. Nell'ipotesi che u_t sia generato da un processo autoregressivo del primo ordine i parametri del modello /3.6/ possono essere stimati con il metodo dei minimi quadrati generalizzati (cfr. Barbone, Bodo e Visco (1981)). All'arrivo delle nuove

informazioni decadali relative al mese T+1 il modello può essere utilizzato per ricavare dalla variazione dei flussi decadali la corrispondente variazione per il totale del sistema. Cumulando queste variazioni si può risalire alle consistenze alla fine di ogni mese del totale dei depositi.

Per valutare il rendimento relativo dei due metodi si è preso in considerazione il biennio 1985-86. La tavola 1 riporta, distintamente per aziende piccole e grandi e per le stime ricavate dai primi invii (stime iniziali) e dai secondi invii (stime riviste), quattro indicatori della performance dei due metodi: l'errore medio assoluto, la deviazione standard degli errori, l'errore medio e l'U di Theil nella definizione proposta in Theil (1966), tutti calcolati rispetto alle consistenze di fine periodo dei depositi bancari. Dato che le stime dei depositi sono essenzialmente utilizzate nel corso dell'anno a fini di interpretazione congiunturale, per meglio apprezzare la rilevanza degli errori, i primi due indicatori sono stati espressi anche in termini di errori di stima dei tassi di crescita mensili destagionalizzati e su base annua (8) derivanti dall'errata stima delle consistenze di fine periodo (tav. 2) (9). Inoltre dato che le informazioni decadali sul mese di dicembre risultano, soprattutto per i primi invii, sensibilmente meno affidabili di quelle degli altri mesi, gli indicatori sono stati anche calcolati escludendo gli errori commessi nei mesi di dicembre e di gennaio che risentono appunto di tale minore affidabilità.

La tavola 1 indica che per entrambi i metodi i risultati sono soddisfacenti; in termini percentuali l'errore medio assoluto risulta di alcuni decimi di punto per le stime iniziali e le stime riviste consentono una riduzione dell'errore nell'ordine del 30-40 per cento includendo dicembre e gennaio e del 10-20 per cento escludendo tali mesi. In quasi tutti i casi alla riduzione dell'errore medio corrisponde anche un calo nella deviazione standard, nell'errore medio e nell'U di Theil. In secondo luogo, si può osservare che la qualità delle stime per dicembre

e gennaio è sensibilmente più bassa. L'errore assoluto viene ridotto di circa il 40 per cento se si escludono tali mesi per le stime iniziali e del 20 per cento per quelle riviste. In terzo luogo gli errori risultano più forti per le aziende piccole come ci si poteva attendere data la ridotta rappresentatività di questo gruppo. Si noti tuttavia che per le aziende grandi le stime iniziali risultano solo di poco migliori di quelle delle aziende piccole: la maggiore rappresentatività è infatti in parte compensata dai maggiori errori commessi dalle aziende grandi sui primi invii. E' solo con i secondi invii che le stime delle aziende grandi risultano nettamente migliori. Si osservi infine che, con entrambi i metodi, l'errore medio (F3) risulta essere particolarmente elevato per le stime iniziali delle aziende piccole; ciò indica la presenza di una distorsione che tende tuttavia a sparire nelle stime riviste.

Il confronto tra i due metodi è leggermente favorevole al primo; per le aziende piccole il riproporzionamento produce un errore medio assoluto inferiore di circa il 10 per cento tanto per le stime iniziali che per quelle riviste. Per le banche grandi, in un caso (la stima iniziale per le aziende grandi sull'intero periodo), il metodo GLS dà risultati migliori, mentre negli altri casi prevale il metodo del riproporzionamento. Simili conclusioni si raggiungono considerando gli altri indicatori, anche se in alcuni casi le stime ottenute col metodo del riproporzionamento presentano una distorsione più elevata di quella GLS, come indicato dal comportamento dell'errore medio. Questo miglior rendimento delle stime ottenute con riproporzionamento è probabilmente dovuto alla maggiore flessibilità del metodo che tiene conto più rapidamente della dinamica delle quote di mercato delle banche incluse nel campione rispetto a quelle escluse.

Nonostante il discreto rendimento dei metodi presentati gli errori commessi non risultano irrilevanti per l'interpretazione congiunturale degli andamenti del mercato monetario. In termini di tassi di crescita (tav. 2) errori medi dell'ordine di 2-4

Errori nella stima del livello dei depositi bancari (8501-8612) (1)

Aziende di credito piccole

		Riproporzionamento		GLS	
		intero periodo	esclusi dic. e gen.	intero periodo	esclusi dic. e gen.
stime iniziali	F1	0,315	0,203	0,362	0,216
	F2	0,468	0,265	0,589	0,308
	F3	0,213	0,101	0,178	0,0789
	F4	5,31	2,80	6,30	3,12
stime riviste	F1	0,205	0,159	0,221	0,174
	F2	0,294	0,223	0,323	0,262
	F3	0,00683	0,00633	0,0181	0,0395
	F4	2,96	2,19	3,22	2,59

Aziende di credito grandi

		Riproporzionamento		GLS	
		intero periodo	esclusi dic. e gen.	intero periodo	esclusi dic. e gen.
stime iniziali	F1	0,284	0,174	0,263	0,239
	F2	0,468	0,219	0,344	0,304
	F3	0,0408	-0,0731	-0,0364	-0,0615
	F4	4,94	2,24	3,41	3,01
stime riviste	F1	0,0969	0,0961	0,113	0,110
	F2	0,129	0,130	0,149	0,146
	F3	0,00888	0,0219	-0,00201	0,0119
	F4	1,28	1,31	1,47	1,45

Nella tavola sono usati i seguenti indici: F1 errore medio assoluto
 F2 deviazione standard
 F3 errore medio
 F4 U di Theil - (Theil, 1966)

Note:

(1) I valori sono espressi in punti percentuali escluso l'U di Theil che è moltiplicato per 1000.

Errori nella stima dei tassi di crescita dei depositi bancari (8501-8612) (1)

Aziende di credito piccole

	Riproporzionamento		GLS	
	intero periodo	esclusi dic. e gen.	intero periodo	esclusi dic. e gen.
stime iniziali	3,98 (5,94)	2,26 (2,72)	4,61 (6,69)	2,78 (3,40)
stime riviste	2,63 (3,62)	2,09 (2,86)	2,88 (4,15)	2,26 (3,27)

Aziende di credito grandi

	Riproporzionamento		GLS	
	intero periodo	esclusi dic. e gen.	intero periodo	esclusi dic. e gen.
stime iniziali	3,37 (5,51)	1,94 (2,50)	3,40 (4,54)	2,84 (3,75)
stime riviste	1,26 (1,66)	1,23 (1,64)	1,50 (1,99)	1,42 (1,87)

Note:

(1) I dati si riferiscono alla media degli errori in valore assoluto dei tassi percentuali di crescita destagionalizzati su base annua; in parentesi sono riportate le deviazioni standard della distribuzione degli errori.

punti percentuali (e di deviazioni standard di 3-5 punti) determinano un disturbo non trascurabile. Sembra possibile ottenere un miglioramento dei risultati cercando di incorporare nella procedura di stima le informazioni ricavabili dalle caratteristiche dinamiche delle serie dei depositi; l'esistenza di una componente sistematica nell'errore dei metodi presentati (la forte stagionalità e, come vedremo, la correlazione presente in alcuni casi) suggerisce inoltre la necessità di tenere in dovuto conto tali caratteristiche. Per entrambi questi fini, il filtro di Kalman costituisce uno strumento adeguato.

4 - Il filtro di Kalman

Il filtro di Kalman, applicato da oltre un ventennio in ingegneria nell'ambito della Teoria dei sistemi, permette di utilizzare in modo efficiente dati affetti da errore combinandoli con un modello dinamico rappresentato in termini di variabili di stato. In particolare può essere utilizzato per ottenere stime efficienti di una data variabile di cui non sono disponibili i dati definitivi a partire da dati provvisori (cfr. Conrad e Corrado (1979), Engle e Watson (1985), Cuthbertson e Taylor (1985), Bordignon e Trivellato (1986)).

Precisamente, per la variabile in esame, occorre costruire un modello dinamico descritto nella forma seguente:

$$/4.1/ \quad X_t = AX_{t-1} + BZ_t + v_t$$

dove X_t detto vettore delle variabili di stato è il vettore contenente gli ultimi n valori dei dati definitivi di cui i più recenti sono non osservabili

Z_t è un vettore di variabili esogene

v_t è un processo bianco gaussiano di data matrice di covarianza

A, B sono matrici di dimensioni opportune contenenti i parametri stimati del modello.

Svariati modelli usati in econometria e statistica possono essere visti come casi particolari del modello considerato (cfr. Bordignon e Trivellato (1986)); in particolare faremo riferimento in questo articolo ai modelli ARIMA univariati.

Ad esempio se si considera un modello AR(n)

$$x_t = a_1 x_{t-1} + a_2 x_{t-2} + \dots + a_n x_{t-n} + e_t$$

la /4.1/ si riduce a

$$X_t = AX_{t-1} + v_t$$

dove

$$A = \begin{vmatrix} a_1 & a_2 & a_3 & \dots & a_{n-1} & a_n \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \dots & 1 & 0 \end{vmatrix} \quad X_t = \begin{vmatrix} x_t \\ x_{t-1} \\ x_{t-2} \\ \cdot \\ \cdot \\ \cdot \\ x_{t-n+1} \end{vmatrix} \quad v_t = \begin{vmatrix} e_t \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{vmatrix}$$

Oltre al modello dinamico relativo ai dati definitivi descritto mediante la /4.1/ occorre costruire un modello di osservazione che descrive la relazione fra i dati definitivi non noti e le loro stime ottenute in via preliminare a partire dalla osservazione diretta, ma con errore del fenomeno.

Il modello di osservazione è descritto nella forma seguente:

$$/4.2/ \quad Y_t = CX_t + DZ'_t + w_t$$

dove Y_t , detto vettore delle osservazioni, è il vettore contenente gli ultimi p valori delle stime preliminari note

Z'_t è un vettore di variabili deterministiche che permettono di incorporare informazioni addizionali

w_t è un processo bianco gaussiano di data matrice di covarianza che rappresenta gli errori di misura

C, D sono matrici numeriche di dimensioni opportune.

Si supponga di conoscere le matrici numeriche A, B, C, D e di conoscere in senso probabilistico lo stato iniziale del sistema X_1 e i disturbi $v(\cdot)$ e $w(\cdot)$; precisamente si supponga che le variabili casuali $X_1, v_1, v_2, \dots, v_t, w_1, w_2, \dots, w_t$ siano congiuntamente gaussiane a valore atteso nullo e fra di loro indipendenti.

Noti i valori assunti dal vettore delle osservazioni $Y(\cdot)$ fino all'istante t si tratta di determinare lo stato del sistema all'istante t .

Si dimostra (cfr. Gelb (1974)) che il previsore ottimo \hat{X}_t per il sistema dinamico /4.1/-/4.2/ considerato è ottenuto dalle seguenti relazioni che prendono complessivamente il nome di filtro di Kalman:

$$/4.3/ \quad \hat{X}_t^- = A \hat{X}_{t-1} + B Z_t$$

$$/4.4/ \quad P_t^- = A P_{t-1} A' + V$$

$$/4.5/ \quad M_t = P_t^- C' (C P_t^- C' + W)^{-1}$$

$$/4.6/ \quad \hat{Y}_t = C \hat{X}_t^- + D Z_t'$$

$$/4.7/ \quad e_t = Y_t - \hat{Y}_t$$

$$/4.8/ \quad \hat{X}_t = \hat{X}_t^- + M_t e_t$$

$$/4.9/ \quad P_t = P_t^- - M_t C P_t^-$$

dove

$$/4.10/ \quad V = \text{Cov}(v_t)$$

$$/4.11/ \quad W = \text{Cov}(w_t)$$

P_t è la matrice di covarianza dell'innovazione sullo stato all'istante corrente t :

$$/4.12/ \quad P_t = \text{Cov} (X_t - \hat{X}_t)$$

e viene calcolata iterativamente a partire da P_{t-1} mediante le equazioni /4.4/, /4.5/, /4.9/; il valore di $P(\cdot)$ all'istante iniziale è assunto come:

$$/4.13/ \quad P_1 = \text{Cov} (X_1)$$

Costruito il valore di $P(\cdot)$ all'istante corrente è possibile calcolare tramite le equazioni /4.3/, /4.6/, /4.7/, /4.8/ la previsione dello stato del sistema dinamico (4.1, 4.2) considerato. I vettori \hat{X}_t^- e P_t^- sono vettori di variabili ausiliarie che rappresentano la stima dello stato e della relativa matrice di covarianza immediatamente prima dell'istante t quando ancora non è disponibile l'osservazione Y_t .

Senza entrare nei dettagli della dimostrazione del teorema di Kalman ora riportato si può cercare di coglierne il meccanismo con alcune semplici considerazioni.

Il filtro di Kalman combina le informazioni a priori date dal modello dinamico /4.1/ con le osservazioni $Y(\cdot)$ che sono informazioni a posteriori. Si supponga di conoscere \hat{X}_{t-1} , di voler ricavare X_t e di avere a disposizione solo le osservazioni passate fino all'istante $t-1$ (cioè Y_t incognito). In tale caso non si può fare altro che basarsi sulle informazioni a priori; poiché $v(\cdot)$ è bianco non si può pensare di ricavare il suo valore attuale noti quelli passati ed essendo additivo e a media nulla, la stima ottenuta con la sola informazione a priori è data dalla relazione /4.3/.

Se in realtà è conosciuto anche Y_t , occorre modificare la relazione che calcola \hat{X}_t per tener conto di questa informazione a posteriori. Il processo con cui il filtro di Kalman effettua questa correzione è il seguente: l'equazione /4.6/ (modello di osservazione) consente di ricavare una stima del vettore di os-

servazione \hat{Y}_t ; dalla differenza tra quest'ultima e il vettore di osservazione Y_t , si ottiene una stima della innovazione occorsa nel periodo (equazione /4.7/); la stima a priori \hat{X}_t^- viene rivista (equazione /4.8/) addizionando un termine dato dal prodotto dell'innovazione e_t per un "peso" M_t detto guadagno del filtro di Kalman dato dalla relazione /4.5/; questa individua l'attendibilità relativa del modello dinamico e delle osservazioni. La matrice dei pesi M_t dipende infatti dalla matrice di covarianza V dell'errore $v(.)$ del modello dinamico /4.1/ tramite le equazioni /4.4/ e /4.9/, che definiscono la matrice di covarianza dell'innovazione sullo stato, e dalla matrice di covarianza W dell'errore $w(.)$ del modello di osservazione /4.2/.

La ragionevolezza del termine $M_t e_t$ può essere vista con i seguenti esempi:

- se l'innovazione e_t è nulla vuole dire che si è determinato esattamente X_t e quindi non occorre aggiungere alcun termine;
- se l'innovazione e_t è diversa da zero ma le osservazioni sono poco attendibili si aggiunge un termine $M_t e_t$ in cui M_t approssima la matrice nulla e si attribuisce quindi scarsa importanza all'innovazione e_t ;
- se l'innovazione e_t è diversa da zero ma il modello dinamico è poco attendibile M_t , nel caso particolare in cui C è quadrato e invertibile, tende ad approssimare C^{-1} e di conseguenza si tende ad approssimare \hat{X}_t con $C^{-1} Y_t + C^{-1} D Z_t'$, cioè con la previsione ottenuta "invertendo" il modello di osservazione /4.2/. A quest'ultima soluzione si può giungere anche nel caso in cui C non sia invertibile ma i passaggi analitici non sono semplici da derivare.

5 - La procedura utilizzata per l'applicazione del filtro di Kalman

In questo paragrafo si descrive la procedura seguita per ot-

tenere, attraverso il filtro di Kalman, una stima "ottimale" della consistenza dei depositi al tempo $t-1$ e al tempo t utilizzando come input le stime iniziali e quelle riviste del flusso dei depositi nei due periodi ottenute con il metodo del riproporzionamento (10). Più precisamente, cinque giorni lavorativi dopo la fine del mese t risultano disponibili, come si è detto, la consistenza dei depositi fino al mese $t-2$, una stima rivista del flusso dei depositi nel mese $t-1$ (V_{t-1}^K) e una stima iniziale del flusso dei depositi nel mese t (V_t^H). Partendo da questi dati la procedura seguita è descritta nella figura 3.

Addizionando alla consistenza R_{t-2} il flusso "rivisto" V_{t-1}^K si ottiene una stima della consistenza dei depositi K_{t-1} che contiene l'innovazione del processo; si ipotizzi che la relazione tra R_{t-1} e K_{t-1} sia:

$$/5.1/ \quad R_{t-1} = K_{t-1} + w_t$$

dove w_t è un processo bianco con deviazione standard pari a σ_w .

La /5.1/ costituisce il più semplice modello di osservazione per l'applicazione del filtro di Kalman. Completando la /5.1/ con un modello ARIMA su R_t :

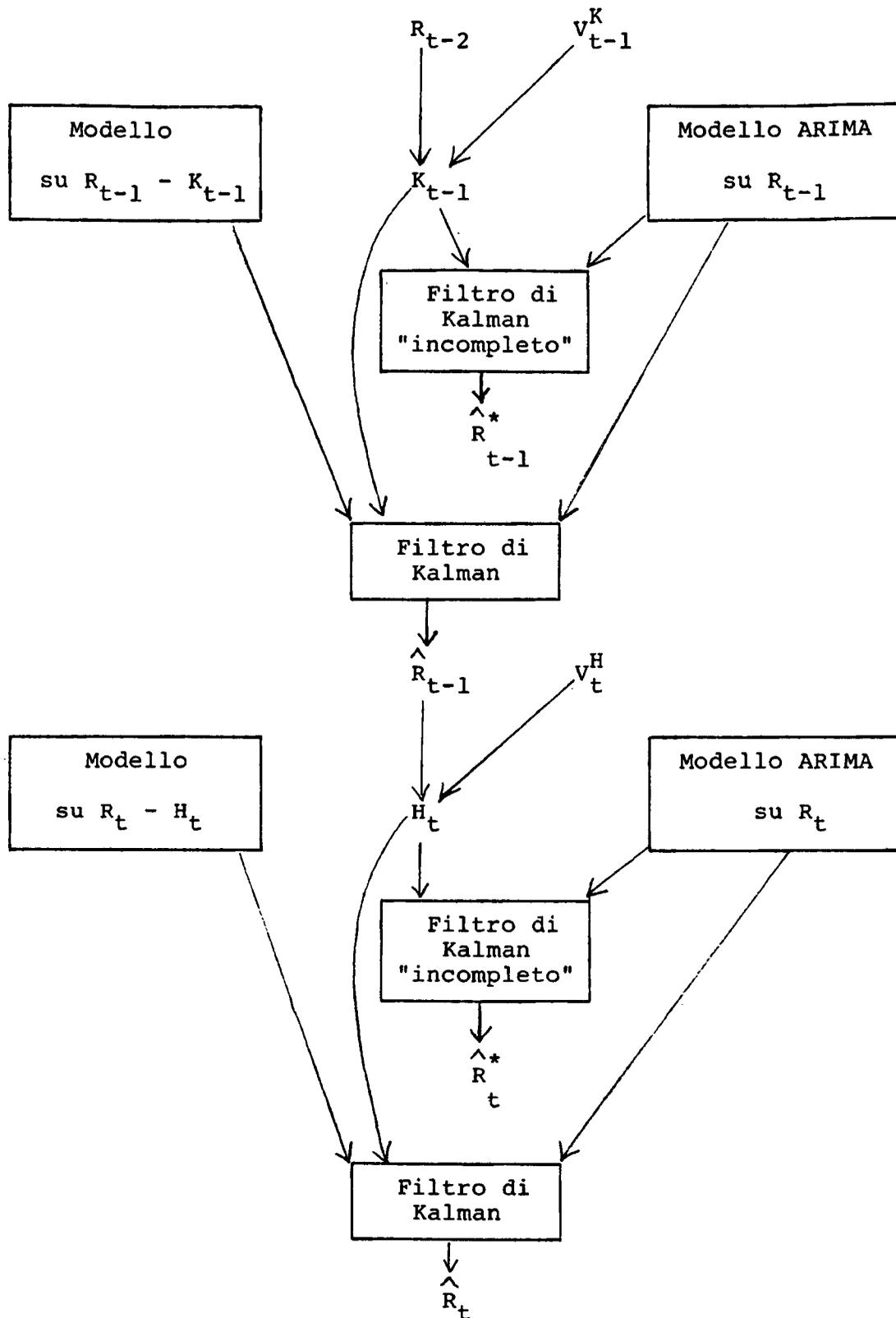
$$/5.2/ \quad \Phi(B) \nabla^d \nabla_s^D R_t = \Psi(B) v_t$$

si ottiene dall'applicazione del filtro di Kalman una stima (ottimale se valgono le /5.1/ e /5.2/) \hat{R}_{t-1}^* ; questa stima è stata definita nelle tavole dei risultati come ottenuta dal filtro di Kalman "incompleto".

L'incompletezza deriva dal non aver ipotizzato una più complessa relazione tra R_{t-1} e K_{t-1} . In realtà è probabile che l'errore di misura $R_{t-1} - K_{t-1}$ non risulti bianco. Si è infatti innanzitutto osservato che la "qualità" delle segnalazioni decedali è inferiore nel mese di dicembre per problemi connessi con la

Fig. 3

Schema di applicazione del filtro di Kalman



chiusura dei bilanci bancari; ciò rende meno affidabili le stime del flusso dei depositi nei mesi di dicembre e gennaio. Anche se questo fenomeno stagionale si è andato riducendo nel tempo sembra necessario tenerne conto. E' inoltre necessario tener conto della possibilità che l'errore di misura non risulti a media zero o sia autocorrelato. Nell'applicazione del filtro di Kalman "completo" la /5.1/ è stata perciò sostituita da un modello di osservazione espresso in forma più generale:

$$\begin{aligned} /5.3/ \quad R_{t-1} - K_{t-1} &= c_0 + c_1(R_{t-1} - K_{t-1})_{-1} + c_2(R_{t-1} - K_{t-1})_{-12} \\ &+ c_3(R_{t-1} - K_{t-1})_{-12} \times \text{DUDIC} + c_4(R_{t-1} - K_{t-1})_{-12} \times \text{DUGEN} \\ &+ u_t \end{aligned}$$

$$/5.4/ \quad u_t = \rho u_{t-1} + w_t$$

Nella /5.3/ le variabili DUDIC e DUGEN sono dummies stagionali che assumono valore unitario rispettivamente a dicembre e gennaio; esse sono state inserite per tener conto della possibile diversa correlazione stagionale che si verifica a dicembre o gennaio, mesi, come si è detto, anomali rispetto agli altri. Come indicato nella figura 3, l'applicazione del filtro di Kalman al sistema formato dalle /5.2/ e /5.3/ consente la derivazione di una stima ottimale della consistenza dei depositi al tempo $t-1(\hat{R}_{t-1})$.

Data la stima "con errore" del flusso dei depositi V_t^H , basato sui primi invii decadali e sul metodo del riproporzionamento, si ottiene una stima "con errore" della consistenza dei depositi al tempo $t(H_t)$. La procedura descritta in precedenza per K_{t-1} , è stata quindi ripetuta per H_t ottenendo una stima sub-ottimale \hat{R}_t^* (filtro di Kalman "incompleto") e una ottimale \hat{R}_t (filtro di Kalman "completo") della consistenza dei depositi al tempo t . L'applicazione del filtro di Kalman incompleto è stata basata sul modello di osservazione:

$$/5.5/ \quad R_t = H_t + w_t$$

Quella del filtro di Kalman completo è basata sul modello:

$$\begin{aligned} /5.6/ \quad R_t - H_t &= d_0 + d_1(R_t - H_t)_{-1} + d_2(R_t - H_t)_{-12} \\ &+ d_3(R_t - H_t)_{-12} \times \text{DUDIC} + d_4(R_t - H_t)_{-12} \times \text{DUGEN} \\ &+ u_t \end{aligned}$$

con u_t modellato come nella /5.4/.

La procedura descritta è stata applicata distintamente ai due gruppi di aziende di credito piccole e grandi; ciò consente di valutare il guadagno connesso con l'uso del filtro di Kalman in due situazioni sensibilmente differenti per grado di affidabilità delle stime preliminari.

6 - Risultati della applicazione del filtro di Kalman

Nell'analisi empirica la stima dei parametri dei modelli ARIMA e dei modelli di misurazione è stata effettuata utilizzando i dati disponibili fino alla fine del 1984 (11). La valutazione dei risultati dell'applicazione del filtro di Kalman alla serie dei depositi bancari è stata effettuata considerando gli errori di previsione compiuti applicando la procedura descritta nel paragrafo 5 sul periodo 1985-86. Ciò consente di valutare il guadagno implicito nell'uso del filtro di Kalman utilizzando unicamente le informazioni disponibili nel momento in cui la stima si riferisce.

Le tavole 3 e 4 riportano, per i due gruppi di aziende, le stime dei parametri del modello ARIMA per R_t e dei due modelli sugli errori di misurazione $R_{t-1} - K_{t-1}$ e $R_t - H_t$, insieme con la stima delle varianze dei tre processi utilizzati per l'applicazione.

segue Tav. 3

4. Modello di misurazione "incompleto" per $r_t - h_t$

$$r_t = h_t + w_t$$
$$\sigma_w^2 = 1,0929 \times 10^{-5} \quad (2)$$

5. Modello di misurazione completo per $r_t - h_t$ (OLS: 8204-8412) (4)(5)

$$r_t - h_t = 0,267(r_t - h_t)_{-12} + 0,370(r_t - h_t)_{-12} \times \text{DUDIC} + w_t$$

$(2,35) \qquad \qquad \qquad (3,16)$

$$\sigma_w^2 = 4,1634 \times 10^{-6}$$

R quadro = 0,966

Durbin-Watson = 2,10

MLM1 = 0,045 (83,5)

MLM12 = 0,490 (49,8)

segue Tav. 4

4. Modello di misurazione "incompleto" per $r_t - h_t$

$$r_t = h_t + w_t$$

$$\sigma_w^2 = 1,621 \times 10^{-5} \quad (2)$$

5. Modello di misurazione completo per $r_t - h_t$ (OLS: 8204-8412) (4)

$$r_t - h_t = 0,891(r_t - h_t)_{-12} \times \text{DUDIC} + w_t$$

(12,54)

$$\sigma_w^2 = 1,906 \times 10^{-5}$$

$$R \text{ quadro} = 0,817$$

$$\text{Durbin-Watson} = 2,02$$

$$\text{MLM1} = 0,123 \quad (73,0)$$

$$\text{MLM12} = 0,000 \quad (99,9)$$

Note alle tavv. 3 e 4

- (1) Nelle tavole 3 e 4 è stata adottata la seguente notazione: $r_t = \log(R_t)$, $h_t = \log(H_t)$, $k_t = \log(K_t)$. Oltre all' R quadro e al Durbin-Watson le tavole riportano, per i modelli di misurazione, il Modified Lagrange Multiplier Test per l'auto-correlazione al primo (MLM1) e al dodicesimo (MLM12) ordine. Per questi ultimi due tests, oltre al valore della statistica F, viene riportata in parentesi la quota percentuale della funzione di densità della distribuzione F alla destra di tale valore per i gradi di libertà esistenti. La stessa informazione è fornita per i tests di Ljung-Box utilizzati per i modelli ARIMA. Sono indicati in parentesi, al di sotto dei coefficienti delle equazioni, i t di Student.
- (2) Queste varianze sono calcolate escludendo gli errori commessi nei mesi di dicembre e di gennaio, mesi in cui la varianza degli errori risultava nel periodo di stima molto più elevata di quella degli altri mesi. Nell'ultimo biennio la qualità delle segnalazioni decadali a dicembre e a gennaio è notevolmente migliorata anche se resta inferiore a quella degli altri mesi.
- (3) L'equazione è stata stimata includendo due dummies puntuali per il gennaio del 1983 e per il febbraio del 1984.
- (4) Il periodo di stima è stato ridotto perché la variabile h_t , in base alle stime di r_{t-1} ottenute con il filtro di Kalman, risulta disponibile per un periodo più breve di k_t . Per omogeneità anche la varianza sui processi ARIMA sono state ricalcolate sul medesimo intervallo; per le aziende piccole si ha $\sigma_v^2 = 4,013 \times 10^{-5}$, per quelle grandi $\sigma_v^2 = 9,543 \times 10^{-5}$.
- (5) L'equazione è stata stimata includendo due dummies puntuali per il dicembre del 1982 e per il febbraio del 1983; si è inoltre introdotta una dummy per il trimestre febbraio-aprile del 1984.
- (6) L'equazione è stata stimata includendo tre dummies puntuali al gennaio e febbraio del 1982 e a gennaio del 1983.

Errori nella stima del livello dei depositi bancari (8501-8612) (1)

Aziende di credito piccole

		Riproporzionamento		Filtro di Kalman incompleto		Filtro di Kalman	
		l'intero periodo esclusi dic. e gen. l'intero periodo esclusi dic. e gen. l'intero periodo esclusi dic. e gen.					
stime iniziali	F1	0,315	0,203	0,260	0,182	0,203	0,176
	F2	0,468	0,265	0,336	0,243	0,272	0,230
	F3	0,213	0,101	0,158	0,101	0,0415	0,0587
	F4	5,31	2,80	3,76	2,55	2,69	2,34
stime riviste	F1	0,205	0,159	0,169	0,137	0,155	0,135
	F2	0,234	0,223	0,243	0,196	0,209	0,188
	F3	0,00683	0,00633	0,00714	0,00557	0,00831	-0,0144
	F4	2,88	2,19	2,37	1,92	2,87	1,86

Aziende di credito grandi

		Riproporzionamento		Filtro di Kalman incompleto		Filtro di Kalman	
		l'intero periodo esclusi dic. e gen. l'intero periodo esclusi dic. e gen. l'intero periodo esclusi dic. e gen.					
stime iniziali	F1	0,284	0,174	0,253	0,152	0,207	0,152
	F2	0,468	0,219	0,386	0,194	0,279	0,194
	F3	0,041	-0,073	0,030	-0,059	-0,092	-0,058
	F4	4,94	2,24	4,04	1,95	2,91	1,95
- stime riviste	F1	0,097	0,096	0,090	0,089	0,096	0,088
	F2	0,129	0,130	0,122	0,121	0,131	0,125
	F3	0,009	0,022	0,008	0,022	0,007	0,018
	F4	1,25	1,31	1,18	1,22	1,26	1,25

Nella tavola sono usati i seguenti indici: F1 errore medio assoluto
 F2 deviazione standard
 F3 errore medio
 F4 U di Theil - (Theil, 1966)

Note:

(1) I valori sono espressi in punti percentuali escluso l'U di Theil che è moltiplicato per 1000.

Errori nella stima dei tassi di crescita dei depositi bancari (8501-8612) (1)

Aziende di credito piccole

	Riproporzionamento	Filtro di Kalman incompleto	Filtro di Kalman
	intero periodo esclusi dic. e gen.	intero periodo esclusi dic. e gen.	intero periodo esclusi dic. e gen.
stime iniziali	3,98 (5,94)	3,27 (4,77)	2,48 (4,11)
	2,26 (2,72)	1,91 (2,56)	1,86 (2,32)
stime riviste	2,63 (3,62)	2,19 (3,09)	1,99 (2,62)
	2,09 (2,86)	1,79 (2,50)	1,77 (2,42)

Aziende di credito grandi

	Riproporzionamento	Filtro di Kalman incompleto	Filtro di Kalman
	intero periodo esclusi dic. e gen.	intero periodo esclusi dic. e gen.	intero periodo esclusi dic. e gen.
stime iniziali	3,37 (5,51)	2,74 (4,38)	2,23 (3,03)
	1,94 (2,50)	1,68 (2,19)	1,68 (2,18)
stime riviste	1,26 (1,66)	1,16 (1,54)	1,23 (1,66)
	1,23 (1,64)	1,14 (1,53)	1,12 (1,58)

Note:

(1) I dati si riferiscono alla media degli errori in valore assoluto dei tassi percentuali di crescita destagionalizzati su base annua; in parentesi sono riportate le deviazioni standard della distribuzione degli errori.

Variazione dell'errore nella stima dei livelli dei depositi bancari (8501-8612) (1)

Aziende di credito piccole

	Filtro di Kalman incompleto		Filtro di Kalman	
	intero periodo	esclusi dic. e gen.	intero periodo	esclusi dic. e gen.
stime iniziali	-17,5 (-28,2)	-10,3 (-8,3)	-35,5 (-41,9)	-13,3 (-12,5)
stime riviste	-17,6 (-17,3)	-13,8 (-12,1)	-24,4 (-28,9)	-15,1 (-15,7)

Aziende di credito grandi

	Filtro di Kalman incompleto		Filtro di Kalman	
	intero periodo	esclusi dic. e gen.	intero periodo	esclusi dic. e gen.
stime iniziali	-10,9 (-17,5)	-12,6 (-11,4)	-27,1 (-40,4)	-12,6 (-11,4)
stime riviste	-7,2 (-5,4)	-7,3 (-6,9)	-1,0 (+1,6)	-8,3 (-3,8)

Note:

(1) I dati si riferiscono alla riduzione percentuale dell'errore medio assoluto rispetto al metodo del riproporzionamento; in parentesi è riportata la riduzione della deviazione standard della distribuzione degli errori.

Variazione dell'errore nella stima dei tassi di crescita dei depositi bancari (8501-8612) (1)

Aziende di credito piccole

	Filtro di Kalman incompleto		Filtro di Kalman	
	intero periodo	esclusi dic. e gen.	intero periodo	esclusi dic. e gen.
stime iniziali	-17,8 (-19,7)	-15,5 (-5,9)	-37,7 (-30,1)	-17,7 (-14,7)
stime riviste	-16,7 (-14,6)	-14,4 (-12,6)	-24,3 (-27,6)	-15,3 (-15,4)

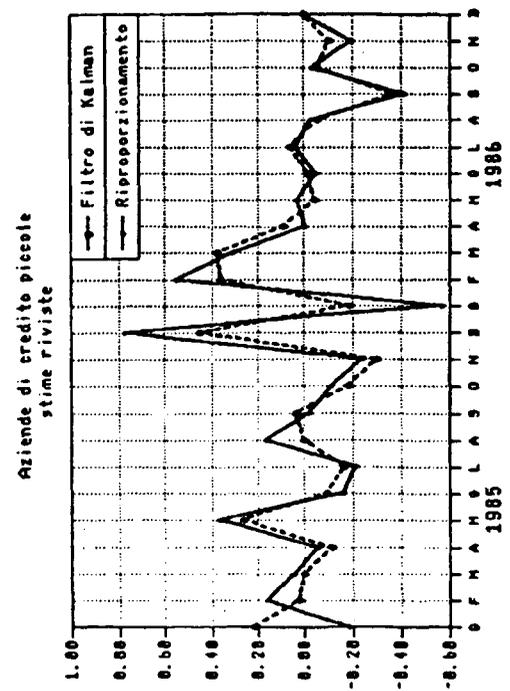
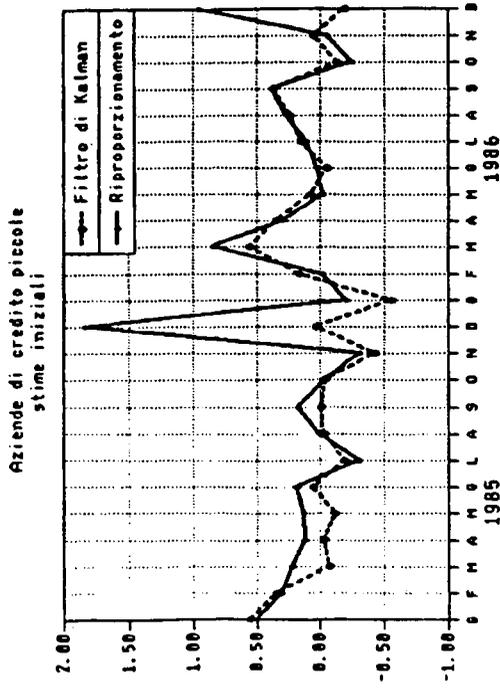
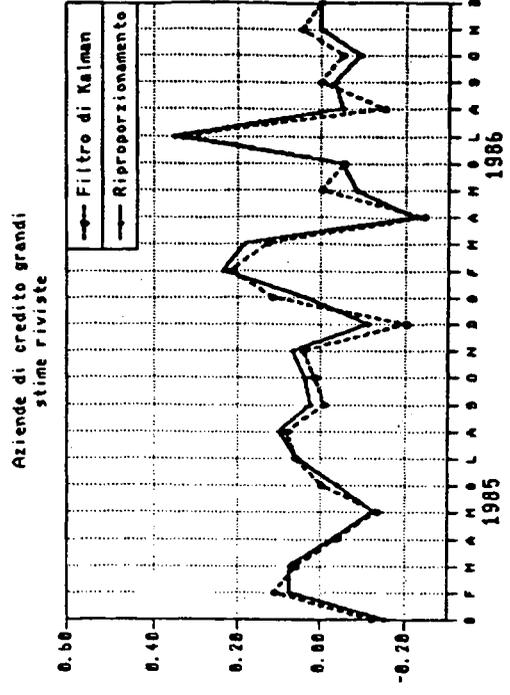
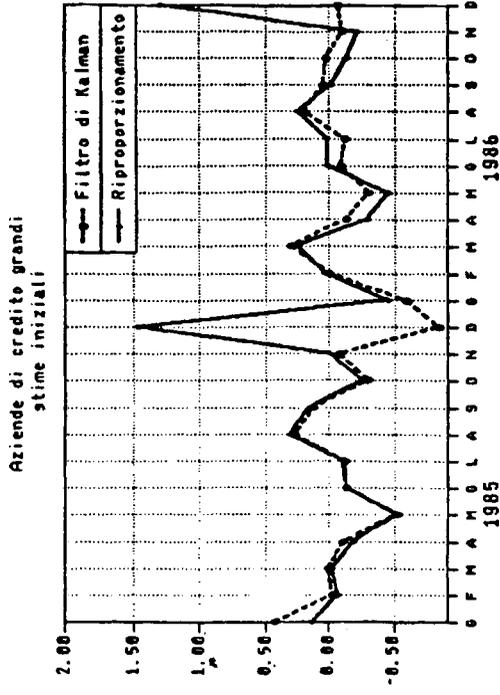
Aziende di credito grandi

	Filtro di Kalman incompleto		Filtro di Kalman	
	intero periodo	esclusi dic. e gen.	intero periodo	esclusi dic. e gen.
stime iniziali	-18,7 (-20,5)	-13,4 (-12,4)	-33,8 (-44,6)	-13,4 (-12,8)
stime riviste	-7,9 (-7,2)	-7,3 (-6,7)	-2,4 (0,0)	-8,9 (-3,7)

Note:

(1) I dati si riferiscono alla riduzione dell'errore medio assoluto rispetto al metodo del riproporzionamento; in parentesi è riportata la riduzione della deviazione standard della distribuzione degli errori.

Errori percentuali nella stima del livello dei depositi bancari (8501-8612)



cazione del filtro di Kalman. Vengono anche riportate le varianze degli errori stessi utilizzate per l'applicazione del filtro di Kalman incompleto.

I risultati dell'applicazione del filtro di Kalman sono riportati nelle tavole 5-8 e nella figura 4. La tavola 5 riporta i quattro indicatori di performance già utilizzati nella tavola 1 calcolati sugli errori nella stima del livello dei depositi bancari per il metodo del riproporzionamento, per il filtro di Kalman incompleto e per quello completo. La tavola 6, con riferimento agli stessi metodi e ai primi due indici, è relativa agli errori nel calcolo dei tassi di crescita mensili destagionalizzati e annualizzati. Per comodità di lettura, le tavole 7 e 8 riportano, rispettivamente per i livelli e per i tassi di crescita, le variazioni percentuali degli errori medi assoluti e dei relativi errori standard che si ottengono applicando il filtro di Kalman incompleto e quello completo rispetto al metodo del riproporzionamento. Infine la figura 4 consente un confronto visivo tra gli errori nella stima delle consistenze ottenute dal riproporzionamento e dal filtro di Kalman.

Per quanto riguarda le aziende di credito piccole, tanto per le stime iniziali che per quelle riviste, la riduzione dell'errore medio assoluto sui livelli e sui tassi di crescita appare sensibile (nell'ordine del 15-20 per cento). La riduzione dell'errore è naturalmente più sensibile se si includono i mesi di dicembre e gennaio per i quali il metodo del riproporzionamento risulta meno efficace: la più forte riduzione dell'errore medio assoluto si ha sulle stime iniziali calcolate per l'intero periodo con un calo del 36 per cento sui livelli e del 38 per cento sulle variazioni percentuali. Occorre rilevare che il miglioramento è sensibile sia per le stime iniziali che per quelle riviste e che anche la dispersione degli errori subisce una riduzione di entità sostanziale. Anche la forte distorsione presente nelle stime iniziali ottenute con il metodo del riproporzionamento si riduce sensibilmente: l'indice F3 passa da 0,213 a

0,042 per l'intero periodo e da 0,101 a 0,059 escludendo dicembre e gennaio. Per quanto riguarda il guadagno relativo tra filtro di Kalman incompleto e completo, sia per le stime iniziali che per quelle finali la modellazione dell'errore sulle stime preliminari consente una sensibile riduzione dell'errore medio assoluto e della sua dispersione. Il guadagno è particolarmente forte sui mesi di dicembre e gennaio, mesi per i quali la componente sistematica dell'errore delle stime preliminari risulta più facilmente identificabile.

Nel complesso il guadagno dell'applicazione del filtro di Kalman appare soddisfacente, anche se inferiore a quello di altri lavori (per esempio, Bordignon e Trivellato (1986) e Conrad e Corrado (1979)); in parte ciò può essere dovuto al fatto che il rendimento del filtro di Kalman è valutato interamente sul periodo successivo a quello per cui le stime dei modelli ARIMA e i modelli di osservazione sono stati effettuati. Questo fattore è particolarmente importante se si considera che sull'ultimo biennio la qualità delle statistiche decadalì è rapidamente migliorata e, probabilmente, le proprietà statistiche delle serie risultano in parte alterate rispetto al passato. E' inoltre da tenere in conto la modesta entità dell'errore compiuto dalle stime preliminari (nell'ordine del 2-3 per mille in termini di livelli) che rende difficili ulteriori miglioramenti.

L'importanza di questo fattore è provata dai risultati dell'applicazione del filtro di Kalman ai dati delle aziende di credito grandi i cui errori sulle stime preliminari risultano più bassi (anche del 50 per cento sulle stime riviste) di quelli delle aziende piccole: risultano corrispondentemente minori i vantaggi ottenuti dall'applicazione del filtro di Kalman. La riduzione dell'errore medio ottenuto con il filtro di Kalman è dell'ordine dell'8-10 per cento anche se in un caso, quello dell'errore sui primi invii calcolati per l'intero periodo con il filtro di Kalman completo, si ottiene una riduzione dell'errore del 27 per cento sui livelli e del 34 per cento sui tassi

di crescita ed un calo della dispersione degli errori rispettivamente del 40 e del 45 per cento. E' interessante osservare che, nel caso delle stime iniziali, quelle ricavate dal filtro di Kalman presentano una distorsione maggiore di quelle ottenute con il metodo del riproporzionamento: l'errore medio (F3) passa da 0,041 a -0,082. Il peggioramento è tuttavia solo apparente: la figura 4 mette in luce che la minor distorsione ottenuta col primo metodo deriva unicamente dai forti errori commessi nei mesi di dicembre che sono di segno opposto rispetto a quelli osservati negli altri mesi. Questa circostanza è di scarsa utilità per un metodo di previsione dell'andamento congiunturale della raccolta bancaria. Escludendo dicembre, l'utilizzo del filtro di Kalman consente, anche per le grandi aziende, una riduzione della distorsione, anche se di importo inferiore a quella ottenuta nel caso delle piccole aziende.

Occorre sottolineare che per le aziende grandi l'applicazione del filtro di Kalman completo determina benefici più contenuti rispetto a quello del filtro incompleto: in particolare, nel caso delle stime riviste valutate sull'intero periodo si determina persino un peggioramento delle stime (da -7,2 a -1 sui livelli e da -7,9 a -2,4 sulle variazioni percentuali; tavv. 7 e 8). Ciò è probabilmente dovuto al più accentuato miglioramento nelle statistiche decadali delle grandi banche sulle quali nell'ultimo biennio si sono accentuati gli sforzi: questo miglioramento rende difficoltosa la modellazione dell'errore delle stime preliminari sulla base dell'esperienza passata.

Sono state finora descritte delle procedure che permettono di ottenere stime dei dati definitivi a partire dai dati del campione decadale relativi alla terza decade di ogni mese.

In linea di principio tuttavia si potrebbe cercare di utilizzare, per la previsione dello stock dei depositi alla fine del mese corrente, solo i dati del campione decadale relativi alla prima e/o la seconda decade.

Con questo obiettivo, e per i soliti due gruppi di aziende, sono stati prodotti modelli ARIMA decadali fino alla fine del 1984; modelli e relativi parametri stimati sono riportati nella tavola 9.

A partire dai modelli ARIMA si sono costruite previsioni ad una decade per l'intero biennio 1985-86 utilizzando come dati contemporanei i dati di primo invio delle seconde decadi e i dati di secondo invio delle decadi precedenti. Le previsioni ottenute, relative alle terze decadi, tradotte in termini di flussi mensili, sono state introdotte nella procedura di riproporzionamento descritta nel paragrafo 3 utilizzando i dati riproporzionati come input della applicazione del filtro di Kalman descritta nel paragrafo 5.

A questo punto l'unica differenza rispetto al caso trattato in precedenza consiste nell'utilizzo come input di una nuova serie mensile H_t^* di stime preliminari al posto di H_t e quindi di una nuova serie di errori di misurazione $R_t - H_t^*$; il modello costruito sulla serie degli errori e i parametri stimati, insieme alla varianza del processo utilizzato per l'applicazione del filtro di Kalman, sono riportati nella tavola 10. I risultati dell'applicazione del filtro di Kalman sono riportati nelle tavole 11 e 12, come al solito relative agli errori rispettivamente nei livelli e nei tassi di crescita.

Come termine di confronto, le tavole riportano gli errori compiuti utilizzando un metodo particolarmente semplice: il dato dei depositi di fine mese è stato ottenuto applicando alla consistenza dei depositi al tempo t_{-12} il tasso di crescita a 36 decadi rilevato nella seconda decade di ogni mese corretto per la differenza tra i tassi di crescita del totale del sistema e del campione decadale all'ultimo dato per cui è disponibile il confronto.

Sia per le aziende di credito piccole che per quelle grandi

Modelli ARIMA decadali dei depositi bancari (1)

1. Aziende di credito piccole

$$\nabla\nabla^{36}d_t = 0,224 \nabla\nabla^{36}d_{t-3} - 0,402\nabla\nabla^{36}d_{t-36} + e_t$$

(3,64) (-2,36)

$$\sigma_e^2 = 3,383 \times 10^{-5}$$

Test di Ljung-Box

Lag	Df	Valore	Prob.
6	4	4,39	35,6
12	10	8,43	58,7
18	16	14,26	58,0

2. Aziende di credito grandi

$$\nabla\nabla^{36}d_t = 0,317\nabla\nabla^{36}d_{t-3} - 0,316\nabla\nabla^{36}d_{t-36} + e_t$$

(3,44) (-2,96)

$$\sigma_e^2 = 4,179 \times 10^{-5}$$

Test di Ljung-Box

Lag	Df	Valore	Prob.
6	4	1,75	78,2
12	10	4,99	89,2
18	16	12,12	73,6

(1) Si è adottata la notazione: $d_t = \log(D_t)$

Modelli di osservazione (1)

1. Aziende di credito piccole

$$r_t - h_t^* = 0,455(r_t - h_t^*)_{-12} \times \text{DUDIC} + w_t$$

(4,16)

$$\sigma_w^2 = 3,15 \times 10^{-5}$$

$$R \text{ quadro} = 0,679$$

$$\text{Durbin-Watson} = 2,10$$

$$\text{MLM1} = 0,154 (70,0)$$

$$\text{MLM12} = 0,026 (87,4)$$

2. Aziende di credito grandi

$$r_t - h_t^* = 0,574(r_t - h_t^*)_{-12} \times \text{DUDIC} + w_t$$

(8,23)

$$\sigma_w^2 = 3,080 \times 10^{-5}$$

$$R \text{ quadro} = 0,701$$

$$\text{Durbin-Watson} = 1,63$$

$$\text{MLM1} = 0,021 (88,8)$$

$$\text{MLM12} = 0,019 (89,3)$$

(1) Si è adottata la notazione: $r_t = \log(R_t)$, $h_t^* = \log(H_t^*)$. Per il significato degli altri simboli si rimanda alla nota 1 relativa alle tavole 3 e 4.

Errori nella stima del livello dei depositi bancari (8501-8612) (1)

Aziende di credito piccole

	Trascinamento del tasso di crescita		Filtro di Kalman	
	intero periodo	esclusi dic. e gen.	intero periodo	esclusi dic. e gen.
stime iniziali				
F1	0,724	0,585	0,407	0,286
F2	0,971	0,780	0,593	0,433
F3	0,0684	0,277	-0,204	-0,0689
F4	9,45	7,95	6,29	4,13

Aziende di credito grandi

	Trascinamento del tasso di crescita		Filtro di Kalman	
	intero periodo	esclusi dic. e gen.	intero periodo	esclusi dic. e gen.
stime iniziali				
F1	0,824	0,777	0,468	0,357
F2	0,961	0,926	0,647	0,445
F3	-0,0578	0,000258	0,0381	0,0169
F4	9,51	9,03	6,41	4,32

Nella tavola sono usati i seguenti indici: F1 errore medio assoluto
 F2 deviazione standard
 F3 errore medio
 F4 U di Theil - (Theil, 1966)

Note:

(1) I valori sono espressi in punti percentuali escluso l'U di Theil che è moltiplicato per 1000.

Errori nella stima dei tassi di crescita dei depositi bancari (8501-8612) (1)

Aziende di credito piccole

	Trascinamento del tasso di crescita		Filtro di Kalman	
	intero periodo	esclusi dic. e gen.	intero periodo	esclusi dic. e gen.
stime iniziali	9,77	7,79	5,39	3,86
	(13,27)	(9,89)	(7,67)	(5,65)

Aziende di credito grandi

	Trascinamento del tasso di crescita		Filtro di Kalman	
	intero periodo	esclusi dic. e gen.	intero periodo	esclusi dic. e gen.
stime iniziali	10,06	8,48	5,65	4,34
	(15,32)	(11,46)	(7,97)	(5,63)

Note:

(1) I dati si riferiscono alla media degli errori in valore assoluto dei tassi percentuali di crescita destagionalizzati su base annua; in parentesi sono riportate le deviazioni standard della distribuzione degli errori.

gli errori sui livelli e sui tassi di crescita appaiono sensibilmente maggiori rispetto a quelli ottenuti usando i primi invii delle terze decadi, il che è spiegabile dalla ridotta informazione di cui si fa uso; in ogni caso utilizzando il filtro di Kalman si ottengono miglioramenti dell'ordine del 40-50 per cento rispetto al metodo che fa uso del trascinamento del tasso di crescita; il guadagno è più forte nei mesi di dicembre e gennaio valendo in questo caso le stesse considerazioni fatte in precedenza.

7 - Conclusioni

Il filtro di Kalman costituisce uno strumento con notevoli potenzialità di applicazione ai problemi di previsione affrontati dagli economisti. Esso risulta particolarmente utile nella stima dell'andamento di una variabile la cui dinamica sia regolata da un modello la cui componente deterministica sia nota (o "stimabile") e i cui valori più recenti siano osservabili con errore. E' questo il caso di molte serie economiche per cui siano disponibili dati provvisori.

Questo lavoro ha presentato una applicazione del filtro di Kalman al problema della stima dei depositi bancari utilizzando come indicatore provvisorio le segnalazioni del campione decennale della Banca d'Italia. L'analisi è stata svolta per i gruppi delle aziende di credito "grandi" e "piccole", come definite nel paragrafo 2, per le stime provvisorie "iniziali" e "riviste" ed evidenziando il guadagno dell'errore nei mesi in cui le segnalazioni provvisorie risultano stagionalmente meno precise. L'applicazione del filtro di Kalman ha consentito di ridurre l'errore dei dati provvisori di circa il 15-20 per cento e di ridurre in misura anche superiore la dispersione degli errori. Il guadagno è stato tanto più sensibile nei casi in cui più imprecise risultavano le segnalazioni provvisorie (le piccole banche, le stime iniziali, i mesi di dicembre e gennaio).

E' stato anche rilevato che un guadagno non trascurabile è ottenibile anche dalla più semplice combinazione, attraverso il filtro di Kalman, di un modello di previsione dinamico con un modello di osservazione in cui si ipotizzi che l'errore di osservazione sia bianco. Naturalmente il guadagno risulta maggiore quando, risultando violata tale ipotesi, anche l'errore di osservazione venga opportunamente modellato.

N o t e

- (1) Si veda, per esempio, Friedman (1975).
- (2) Si veda la classificazione dimensionale delle aziende di credito italiane inclusa nella Appendice alla Relazione della Banca d'Italia sul 1986.
- (3) Le aziende di credito che inviano mensilmente le segnalazioni di Vigilanza erano, alla fine del 1986, 285, rispetto alle 1102 aziende in esercizio. Queste aziende, che costituiscono il cosiddetto "campione mensile", rappresentano tuttavia più del 95 per cento del totale dei depositi bancari. Ai fini di analisi congiunturale il campione mensile viene perciò considerato convenzionalmente pari al totale del sistema. Anche in questo lavoro viene adottata questa convenzione.
- (4) Più esattamente le aziende trasmettono, con il ritardo indicato, informazioni relative alla consistenza delle voci di bilancio al 10, al 20 e alla fine del mese. In ogni anno esistono quindi convenzionalmente 36 decadi; la terza decade di ogni mese ha durata variabile a seconda del mese (e, per febbraio, dell'anno) considerato. In caso di festività i dati trasmessi si riferiscono all'ultimo giorno lavorativo precedente quello di riferimento.
- (5) Come si è detto il calcolo, incluso quello dei coefficienti di rappresentatività, viene effettuato separatamente per le aziende grandi e per quelle piccole.
- (6) Naturalmente il criterio di esclusione presentato è arbitrario. Criteri più "oggettivi", ad esempio l'esclusione delle aziende che presentavano tassi di crescita differenti dalla crescita media più di n deviazioni standard risultano in teoria preferibili perchè la crescita dei depositi è fortemente irregolare nel tempo (basta pensare alla fortissima stagionalità del mese di dicembre). D'altro canto i metodi di esclusione basati sulla deviazione standard sono molto sensibili alla presenza di uno o più errori particolarmente forti che distorcono fortemente il calcolo della stessa deviazione standard; ciò richiede il ricalcolo una o più volte di tale deviazione dopo l'esclusione dei dati anomali. Nella pratica quotidiana è questo il metodo seguito, ma per una analisi estesa al passato si è ritenuto che il più semplice criterio di esclusione descritto risultasse sufficiente a individuare gli errori di segnalazione.
- (7) Più esattamente la stima è effettuata, come si è detto, cinque giorni lavorativi dalla fine del mese t . Un metodo equivalente a quello esposto nel testo è naturalmente quello di calcolare H_t come: $R_{t-2} + v_t^{H,2}$ dove $v_t^{H,2}$ è il flusso nei due

mesi terminanti a t calcolato come $D1_t - D2_{t-2}$ e riproporzionato in base a p_t .

- (8) In linea di principio l'analisi degli errori commessi sulle stime dei tassi di crescita mensili destagionalizzati è meno significativo di quella sui livelli; l'aggiustamento stagionale è infatti una procedura "ad hoc" e in quanto tale contiene un elemento di arbitrarietà soprattutto quando, come in questo lavoro, essa è basata su metodi a medie mobili come l'X11ARIMA. Si è tuttavia ritenuto opportuno fornire anche questa "chiave di lettura", limitatamente all'errore medio assoluto e alla deviazione standard, perché, nella pratica quotidiana, l'analisi congiunturale è basata in buona parte sull'utilizzo di tali tassi di crescita.
- (9) Nel calcolare i tassi di crescita sono state naturalmente utilizzate le informazioni migliori disponibili al tempo t . Così, ad esempio, la stima "iniziale" del tasso di crescita relativa al mese t è stata ottenuta rapportando le stime "iniziali" dei depositi alla consistenza "rivista" del mese $t-1$. Si noti inoltre che nella applicazione del metodo GLS i parametri del modello sono ristimati di mese in mese all'arrivo di nuove informazioni disponibili e che, come nel primo metodo, sono state preliminarmente escluse le aziende che presentano nella decade variazioni anomale.
- (10) Si è deciso di utilizzare come stima preliminare per il filtro di Kalman la stima del flusso ottenuto con il metodo del riproporzionamento in quanto superiore a quella ottenuta con il metodo GLS, come descritto nel paragrafo 3.
- (11) In linea di principio risulterebbe ottimale ristimare i parametri dei modelli via via che nuove informazioni diventano disponibili. Per semplicità di programmazione, si è tuttavia preferito operare con parametri fissi, rinviando l'introduzione di parametri variabili a una espansione futura del lavoro.

Bibliografia

- Barbone, L. - Bodo, G. - Visco, I. (1981), Costi e profitti nell'industria in senso stretto: un'analisi su serie trimestrali, Banca d'Italia, "Bollettino", gen.-dic.
- Bordignon, S. - Trivellato, U. (1986), On the Optimal Use of Provisional Data in Forecasting with Dynamic Models, Università di Padova, mimeo.
- Ciampi, C.A. (1987), Mercati finanziari in evoluzione: riflessi per il governo della moneta e del credito, Banca d'Italia, "Bollettino economico", n. 8, feb.
- Conrad, W. - Corrado, C. (1979), Applications of the Kalman filter to revisions in monthly retail sales estimates, "Journal of Economic Dynamics and Control", 1, pp. 177-198.
- Cuthbertson, K. - Taylor, M. (1985), Monetary anticipations and the demand for money: results for the UK using the Kalman Filter, University of Newcastle-upon-Tyne, Dept. of Economics, "Discussion Papers", n. 90.
- Engle, R. - Watson, M. (1985), Applications of Kalman Filtering in Econometrics, Cambridge (Mass.), Harvard University, "Discussion Papers", n. 1187.
- Friedman, B. (1975), Targets, Instruments and Indicators of Monetary Policy, "Journal of Monetary Economics", pp. 443-473.
- Gelb, A. (1974), Applied Optimal Estimation, Cambridge (Mass.), MIT Press.
- Harvey, A.C. (1981), Time Series Models, London, Philip Allan.
- Theil, H. (1966), Applied Economic Forecasting, Rand McNally & Co.

ELENCO DEI PIÙ RECENTI TEMI DI DISCUSSIONE (*)

- n. 82 — *La redditività bancaria in Italia. Problemi metodologici e aspetti empirici*, di P. MARULLO REEDTZ - F. PASSACANTANDO (dicembre 1986).
- n. 83 — *Domanda di lavoro e trasformazione dell'economia del Mezzogiorno: l'esperienza degli ultimi decenni e le prospettive. Il ruolo degli incentivi all'occupazione*, di F. SIRACUSANO - C. TRESOLDI - G. ZEN (dicembre 1986).
- n. 84 — *Interscambio con l'estero e struttura produttiva: elementi per un'analisi integrata*, di P. CASELLI - L. F. SIGNORINI (febbraio 1987).
- n. 85 — *Innovazione finanziaria e attività di merchant banking*, di F. CAPRIGLIONE (marzo 1987).
- n. 86 — *L'indagine sui bilanci delle famiglie italiane nel 1984. Alcune considerazioni sugli errori campionari*, di L. CANNARI (aprile 1987).
- n. 87 — *Aspetti macroeconomici dell'interazione fra sviluppo ed energia*, di R. S. MASERA (aprile 1987).
- n. 88 — *La tassazione dei titoli pubblici in Italia: effetti distributivi e macroeconomici*, di G. GALLI (aprile 1987).
- n. 89 — *Shocks temporanei e aggiustamento dinamico, una interpretazione contrattuale della CIG*, di L. GUISO - D. TERLIZZESE (luglio 1987).
- n. 90 — *Il rientro dell'inflazione: un'analisi con il modello econometrico della Banca d'Italia*, di D. GRESSANI - L. GUISO - I. VISCO (luglio 1987).
- 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 - N. PIETRAFESA (agosto 1987).

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

