BANCA D'ITALIA

Temi di discussione

del Servizio Studi

L'imputazione di informazioni mancanti: una sperimentazione

di Luigi Cannari

Numero 100 - Marzo 1988

BANCA D'ITALIA

Temi di discussione

del Servizio Studi

L'imputazione di informazioni mancanti: una sperimentazione

di Luigi Cannari

Numero 100 - Marzo 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.

RIASSUNTO

Le indagini campionarie sono spesso caratterizzate dal problema dei missing data. Di frequente accade che le unità campionarie non siano in grado di rispondere a taluni quesiti o forniscano informazioni inattendibili, che non superano le fasi di controllo di coerenza dei dati rilevati.

La presenza di missing data può determinare rilevanti distorsioni nella stima di valori medi o di totali, soprattutto quando le percentuali di non rispondenti variano sensibilmente tra i vari sottoinsiemi del campione.

E' pertanto necessario, al fine di migliorare la qualità dell'informazione statistica, sperimentare delle tecniche di imputazione.

Nel presente lavoro tale sperimentazione viene effettuata per la variabile patrimonio in immobili, rilevata nell'indagine sui bilanci delle famiglie, e per le variabili investimento effettivo e previsto, rilevate nell'indagine sulle imprese.

INDICE

1. Introduzione (*)

Nelle indagini campionarie vengono generalmente posti numerosi quesiti a ciascuna delle unità estratte e l'insieme delle informazioni rilevate risulta spesso caratterizzato dal problema dei missing data.

La rilevanza di tale fenomeno è diversificata a seconda del numero e del tipo di variabili oggetto di indagine; il tasso di non risposta è infatti più elevato per le informazioni considerate "riservate" dalle unità campionarie ed è correlato con la dimensione e la struttura del questionario, che influenzano sensibilmente la disponibilità a collaborare del campione.

Il fenomeno è correlato inoltre con le caratteristiche degli individui incaricati di svolgere la rilevazione e con l'accuratezza con la quale il questionario viene presentato e compilato; non sono infatti infrequenti i missing data che derivano da una "domanda non posta" dall'intervistatore o dalla difficoltà, da parte dell'intervistato, di comprendere taluni quesiti o di fornire valutazioni o stime non sempre semplici.

Infine, talune risposte sono poste a missing in seguito alla fase di controllo delle informazioni rilevate, nel caso di incoerenza del contenuto di singoli campi o di campi correlati.

Il problema dei missing data assume una particolare importanza nelle indagini campionarie di tipo quantitativo, quando le variabili oggetto di indagine

---

(*)L'autore desidera ringraziare Gabriella Callini, Piero Falorsi, Daniela Gressani e Achille Lemmi. Utili suggerimenti sono stati tratti dalle osservazioni di un anonimo "referee".

sono calcolate come aggregati di componenti elementari; l'assenza di uno o più dati elementari può creare infatti serie distorsioni nella stima di valori medi o di totali, nel caso che la percentuale delle "non risposte" sia elevata e non si utilizzi alcun metodo di imputazione.

Al fine di valutare quali metodologie, tra quelle frequentemente impiegate, risultano più efficienti per il trattamento dei missing data nelle indagini della Banca d'Italia, sono stati effettuati studi empirici sui dati relativi (a) alla ricerca sul reddito e patrimonio delle famiglie italiane nell'anno 1986 e (b) all'indagine sugli investimenti delle imprese industriali.

I risultati relativi al caso (a), riportati nel presente lavoro (paragrafi 2.1 e 2.2), riguardano la stima del patrimonio in immobili; la scelta di tale variabile è dovuta alla maggiore presenza di missing data rispetto alle altre variabili oggetto di indagine 1/ e all'importanza di tale componente nella stima della ricchezza reale complessiva netta 2/. I risultati relativi al caso (b) riguardano le variabili investimento effettivo e investimento programmato e sono riportati nei paragrafi da 3.1 a 3.3.

Per facilitare il lettore, è riportata nell'appendice "A" una tabella dei simboli utilizzati.

2.1 I metodi utilizzati per l'imputazione e la stima del patrimonio in immobili

I metodi di trattamento dei missing data (in seguito MD) derivano da particolari assunzioni riguardo

le probabilità di non risposta di ciascuna unità campionaria o di gruppi di queste. Nel corso del lavoro si farà riferimento a metodi di correzione basati su una probabilità di non risposta indipendente dal livello della variabile da imputare e correlata con le caratteristiche socio demografiche delle famiglie e con la tipologia dell'immobile. Infatti, mentre da un lato non è ragionevole assumere lo stesso meccanismo probabilistico di non risposta per il complesso del campione, dall'altro risulta complesso modellare tale meccanismo in funzione di valori non osservati (Rubin, 1983). Essendo questo lavoro finalizzato alla valutazione del problema dei MD con particolare riferimento ad una specifica indagine si è reso pertanto necessario analizzare, prima di sperimentare un qualunque metodo di imputazione, le caratteristiche probabilistiche delle mancate risposte.

Nell'indagine sui bilanci delle famiglie italiane (Banca d'Italia, 1987a), il valore degli immobili di proprietà della famiglia risulta "non indicato" nel 5-6 per cento circa dei casi, generalmente a causa della difficoltà, per l'intervistato, a stimare il valore di mercato delle proprietà immobiliari della famiglia.

Tale percentuale è notevolmente differenziata per caratteristiche della famiglia e, in misura minore, per tipologia dell'immobile; si osserva infatti una correlazione positiva con l'età del capofamiglia (in seguito CF) e negativa con il livello di istruzione di quest'ultimo (tavola 1). Per quanto riguarda il tipo di immobile, una maggiore probabilità di non risposta caratterizza i terreni agricoli e non agricoli rispetto alle abitazioni e altri fabbricati; ciò è probabilmente dovuto al minor numero di scambi che si realizzano nel mercato dei terreni con la conseguente difficoltà di disporre di valori di riferimento.

Per condurre l'analisi empirica sulle caratteristiche dei diversi metodi di imputazione si è assunto come popolazione di riferimento il sottoinsieme del campione B.I. relativo al patrimonio immobiliare, escludendo dalle elaborazioni tutti i questionari caratterizzati dalla presenza di missing data nei campi "valore dell'immobile", "tipo di immobile", "titolo di studio" ed "età" del CF, che esauriscono l'insieme delle variabili utilizzate per l'analisi 3/.

Al fine di generare casualmente MD con una struttura simile a quella effettivamente realizzatasi nel 1986, si è ipotizzata una funzione di probabilità di non risposta di tipo logit 4/; tale funzione è stata stimata sulla base delle percentuali di non rispondenti riportate nella tavola 1, utilizzando, come variabili dipendenti, il tipo di immobile, l'età e il livello di istruzione del capofamiglia (modello 1) 5/.

In taluni casi, tuttavia, la probabilità di non risposta può dipendere dal valore della variabile non disponibile (David, Little, Samuhel and Triest, 1983; Little, 1983); ad esempio, la reticenza degli intervistati a fornire informazioni sul reddito o la ricchezza potrebbe essere maggiore per le famiglie più ricche.

Di questo fenomeno non si è tenuto conto nella fase di stima della probabilità di non risposta; tuttavia, per analizzare gli effetti che un comportamento del genere determina nei risultati delle procedure di imputazione, il modello 1 sopra descritto è stato generalizzato introducendo arbitrariamente (con un coefficiente pari a 25×10^{-7}, che incrementa di circa l'uno per cento la probabilità media di non risposta) il "valore dell'immobile" tra le variabili esplicative della probabilità di non risposta (modello 2).

La generazione casuale di MD, utilizzando i

Tav.1

Stima della probabilità di non risposta per tipo di immobile e caratteristiche del capofamiglia

Modalità /	Probabilità di non risposta(1)	Stima della probabilità di non risposta(2)
Età(anni) (3)		
fino a 30	1,0	3,1
30 - 40	4,3	3,4
41 - 50	4,7	4,2
51 - 65	6,0	5,5
oltre 65	7,9	7,3
Titolo di studio(3)		
laurea	4,4	3,2
lic.med.sup.	3,8	3,6
lic.med.inf.	3,1	4,5
lic.element.	6,9	6,2
nessun titolo	11,7	7,6
Condizione professionale(3)		
dipendente	4,7	4,9
autonomo	4,5	4,3
cond.non prof.	7,9	6,5
Tipo di immobile		
Abitazione principale	5,2	4,5
Altra abitazione	4,5	4,0
Altro fabbricato	5,5	4,2
Terreno agricolo	9,7	9,8
Terreno non agricolo	5,9	8,6
Totale	5,7	5,2

(1) Valore calcolato sui dati campionari dell'indagine B.I..
(2) Valori stimati con il modello logit.
(3) Con riferimento al capofamiglia.

modelli 1 e 2, ha consentito di sperimentare alcune procedure di stima del valore medio del patrimonio in immobili; in particolare si è fatto ricorso ai seguenti stimatori:

$$(1)\quad \bar{x}_1 = (1/N) \sum_{i}^{R} x_{Ri}$$

$$(2)\quad \bar{x}_2 = (1/R) \sum_{i}^{R} x_{Ri}$$

$$(3)\quad \bar{x}_3 = (1/N)\left(\sum_{i}^{R} x_{Ri} + (N-R)\,\bar{x}_2\right)$$

$$(4)\quad \bar{x}_4 = (1/N)\left(\sum_{h}^{L}\sum_{i}^{h} x_{Rhi} + \sum_{h}^{L}\sum_{i}^{h}\bar{x}_{Rh}^{(N-R)}\right)$$

dove:

$$\bar{x}_{Rh} = 1/R \sum_{h}^{R}\sum_{i} x_{Rhi}$$

$$(5)\quad \bar{x}_5 = (1/N)\left(\sum_{h}^{L}\sum_{i}^{h} x_{Rhi} + \sum_{h}^{L}\sum_{j}^{h} x_{Rhj}^{(N-R)}\right)$$

dove:

x_{Rhj} è estratto casualmente dall'insieme $(x_{Rhi})_{i=1,R_h}$

$$(6)\quad \bar{x}_6 = 1/N \left(\sum_{i}^{R} x_{Ri} + \sum_{j}^{(N-R)} x''_{NRj} \right)$$

dove:

$$x''_{NRj} = g(Y_j\ P_j)$$

indica che il valore della variabile x per il j-esimo

MD è stimato sulla base del vettore Y di variabili disponibili e della probabilità P di non risposta. Il vettore Y utilizzato per la stima comprende il tipo di immobile, il livello di istruzione e l'età del CF; la funzione g(.) è ipotizzata lineare 6/.

Con riferimento agli stimatori si nota che il primo di questi, denominato zero substitution method (Platek and Gray, 1983), equivale a non prendere alcuna iniziativa per risolvere il problema dei MD nelle componenti di una variabile aggregata; in questo caso pertanto la ricchezza reale complessiva risulterebbe costituita solo dagli immobili per i quali il proprietario è in grado di fornire informazioni sul valore di mercato. La componente ricchezza in immobili è pertanto sottostimata e la sottostima è direttamente correlata con il tasso di non risposta. Per quanto riguarda gli altri stimatori sperimentati è importante precisare che la formula (2) è identica alla (3) e fornisce gli stessi risultati, come è semplice mostrare; con la diversa rappresentazione formale si vuole sottolineare che nel caso (2) i MD sono esclusi dalla elaborazione mentre nel caso (3) questi sono imputati ed assumono come valore la media calcolata su tutto il campione dei rispondenti; ciò non ha effetti sulla stima della media ma modifica il valore stimato della varianza delle osservazioni, come evidenziato nel paragrafo successivo. A differenza del metodo (3), lo stimatore (4) è fondato su un meccanismo di risposta variabile tra i diversi sottoinsiemi del campione. Pertanto il campione complessivo viene ripartito in classi di imputazione utilizzando variabili ausiliarie (ad esempio le caratteristiche socio-demografiche della famiglia e/o la tipologia dell'immobile) e i MD sono imputati sulla base dei valori medi stimati per i rispondenti in ciascuna di tali classi. Utilizzando i metodi (3) e (4) si tende tuttavia a sottostimare il

valore della varianza delle osservazioni, poiché ad ogni MD viene sostituito un valore medio. L'esigenza di ridurre tale distorsione è alla base del metodo (5); in tal caso infatti i MD sono stimati con una estrazione casuale in ciascuna cella di imputazione, al fine di riprodurre la distribuzione osservata delle variabili.

Nel modello (6), infine, i MD sono imputati utilizzando una funzione lineare delle variabili disponibili per ciascun elemento del campione e della propensione a rispondere; ciò al fine di tenere conto dell'eventuale legame esistente tra la probabilità di risposta e il livello della variabile da imputare.

2.2 I principali risultati per la variabile ricchezza in immobili

La replicazione sia del procedimento di generazione casuale dei MD che della stima dei valori medi per le diverse tecniche di trattamento delle mancate risposte (6 valori medi stimati per 10 simulazioni indipendenti) ha consentito di disporre di informazioni sulle caratteristiche degli stimatori; i risultati relativi a ciascun stimatore sono stati sintetizzati calcolandone la media per le 10 simulazioni; quest'ultima è stata confrontata con i valori "veri" calcolati per la popolazione di riferimento.

I risultati sono riportati nella tavola 2; la maggiore distorsione (4,0 per cento) si verifica quando nessuna tecnica di imputazione viene utilizzata per il trattamento dei MD. Tale distorsione risulta superiore per il modello 2, in quanto, per costruzione, le non risposte sono più probabili per gli immobili di valore elevato. Si noti che, quando tutte le unità campionarie sono caratterizzate dalla stessa probabilità di non risposta, la distorsione dello stimatore (1) è in media

Tav.2

Distorsioni relative alle stime del valore medio e del coefficiente di variazione

Metodo	Distorsione percentuale			
	Modello 1 (*)		Modello 2 (**)	
	stima della media	stima del c.v.($)	stima della media	stima del c.v.($)
(1)	-4,0	+4,8	-6,0	+2,4
(2)	+0,7	-0,6	-0,7	-3,4
(3)	+0,7	-2,9	-0,7	-6,3
(4)	+0,4	-2,3	-1,2	-5,5
(5)	-0,1	-0,6	-1,5	-3,2
(6)	+0,0	-1,9	-1,3	-5,1

(*) Probabilità media di non risposta per le simulazioni effettuate: 5 per cento.
(**) Probabilità media di non risposta per le simulazioni effettuate: 5 per cento.
($) Il coefficiente di variazione ((s.q.m./media)x100) ed il valore medio delle osservazioni utilizzate come popolazione di riferimento risultano rispettivamente pari a 89,7 ed a 80.112.

pari alla quota di MD sul totale delle osservazioni; in tal caso infatti

$$E(\bar{x}_1) = 1/N \sum_{i}^{R} E(x_{Ri}) = R/N\ E(x)$$

e la distorsione, cioè la differenza tra il valore medio dello stimatore e la media della popolazione di riferimento (E(x)), è pari a [(R-N)/N] E(x), ovvero proporzionale al tasso di non risposta.

Questo fenomeno non si verifica per le tecniche (2) e (3); queste sono in media corrette quando tutte le unità sono caratterizzate dalla stessa probabilità di non risposta 7/. L'ipotesi di assoluta casualità delle non risposte non è tuttavia suffragata dai dati utilizzati per l'analisi, come evidenziato nel paragrafo precedente; ciò determina, utilizzando la (2) o la (3) una distorsione nella stima del valor medio pari allo 0,7 per cento di quest'ultimo.

Si osservi che lo stimatore (2) (e di conseguenza anche lo stimatore (3)), nel caso di non risposte casuali con probabilità diverse per i vari sottoinsiemi della popolazione di riferimento, è caratterizzato da una distorsione esprimibile in funzione dei valori medi di tali sottoinsiemi. Il valore medio dello stimatore (2) risulta, in queste ipotesi, pari a :

$$E(\bar{x}_2) = E(1/R \sum_{i}^{R} x_{Ri}) = E(1/R \sum_{h}^{L} R_h \bar{x}_{Rh}) =$$

$$= 1/(N(1-\bar{P})) \sum_{h}^{L} (1-P_h) N_h E(x_{hi})$$

Tale valore è pari alla media della popolazione

$(E(x))$ se $(1-P_h)=(1-\bar{P})$ $(h=1,L)$, ovvero quando le probabilità di non risposta sono uguali in ogni sottoinsieme. Infine la distorsione:

$$E(\bar{x}_2)-E(x)$$

è positiva quando la probabilità di non risposta è maggiore per i sottoinsiemi della popolazione caratterizzati da un valore medio inferiore alla media calcolata per il complesso dei dati, come accade per la variabile "valore dell'immobile" nel campione B.I.8/.

Per quanto riguarda gli stimatori (4), (5) e (6), questi sono corretti anche nell'ipotesi che le probabilità di non risposta siano diverse tra i vari sottoinsiemi della popolazione di riferimento. Tali stimatori richiedono tuttavia che i sottoinsiemi caratterizzati da un diverso tasso di non risposta siano correttamente identificati e che le probabilità di non risposta siano uguali per tutte le unità appartenenti allo stesso sottoinsieme. Pertanto, qualora le celle di imputazione (per gli stimatori (4) e (5)) o la funzione g(.) (per lo stimatore (6)) non siano correttamente identificate, potranno verificarsi distorsioni nella stima dei valori medi. A questo proposito si nota che nell'analisi empirica svolta lo stimatore (4) è caratterizzato da un errore in media pari allo 0,4 per cento del valore medio calcolato sui dati di riferimento. Nella costruzione delle celle di imputazione relative a tale stimatore si è infatti tenuto conto solo del tipo di immobile e non dell'età e del titolo di studio del capofamiglia.

Per gli stimatori (5) e (6) le celle di imputazione e la funzione di probabilità di non risposta sono correttamente individuate (per costruzione) e la distorsione percentuale stimata risulta rispettivamente pari a -0,05 e 0,02; in caso di corretta specificazione si ottiene infatti:

$$E(\bar{x}_5)=1/N\left(\sum_h^L \sum_i^{R} E(x_{hi}) + \sum_h^L \sum_i^{N-R} E(x_{hi})\right)=1/N \sum_h^L N_h E(x_{hi})=E(x)$$

$$E(\bar{x}_6)=1/N\left(\sum_i^{R} E(x_{Ri}) + \sum_j^{N-R} E(x''_{NRj})\right)=1/N\left(\sum_i^{R} E(x_i) + \sum_j^{N-R} E(x_j)\right)=E(x)$$

Per quanto riguarda il secondo modello di probabilità di non risposta (modello 2) si è ritenuto interessante calcolare gli stimatori sulla base delle stesse celle di imputazione utilizzate per il modello 1; tale scelta deriva in primo luogo dall'esigenza di analizzare le differenze che intercorrono tra le stime dei valori medi nel caso di specificazione non corretta (delle celle di imputazione), dovuta alla impossibilità di conoscere in che misura la probabilità di non risposta dipende dal valore non indicato della variabile; in secondo luogo è generalmente ignoto e raramente stimabile il legame tra le probabilità di non risposta e la variabile da imputare.

Ad un primo esame della tavola 2 si nota che le tecniche (2) e (3) forniscono i migliori risultati; la modesta distorsione che caratterizza lo stimatore (2) (e di conseguenza (3)) è tuttavia determinata da una somma di componenti di segno contrario: da una parte la distorsione positiva dovuta alla maggiore probabilità di non risposta per i terreni (con un valor medio inferiore al valor medio per il complesso degli immobili), dall'altra una componente negativa, analoga per tutti gli stimatori esaminati, determinata dalla specificazione probabilistica del modello 2.

Si può quindi concludere che solo apparentemente le tecniche (2) e (3) forniscono i risultati migliori nel caso di probabilità di non risposta positivamente correlate con il valore dell'immobile.

Infine, per quanto attiene le stime della variabilità delle informazioni 9/, si osserva che la tecnica (1) sovrastima, nel caso studiato, l'effettivo coefficiente di variazione delle osservazioni (94,0 contro 89,7); non utilizzare alcuna tecnica di imputazione per il calcolo di variabili aggregate equivale infatti ad azzerare la componente elementare non indicata; ciò causa una distorsione non necessariamente negativa nella stima della varianza.

Per quanto riguarda la tecnica (2) la distorsione (-0,6 per cento) relativa al coefficiente di variazione (c.v.) delle osservazioni è determinata dalla diversa probabilità di non risposta che caratterizza i sottoinsiemi del campione distinti per tipo di immobile e caratteristiche del CF; in caso di assoluta casualità (probabilità di non risposta uguali per ogni elemento del campione) delle non risposte la varianza delle osservazioni stimata con la tecnica (2) risulta infatti:

$$S_2^2 = 1/(R-1) \sum_{i}^{R} (x_{Ri} - \bar{x}_2)^2$$

che, in media, è corretta.

A differenza di quanto accade per la tecnica (2), il metodo (3) di imputazione implica una distorsione negativa (c.v. stimato: 87,1; c.v. effettivo: 89,7) nella stima del coefficiente di variazione delle osservazioni, poiché ogni MD viene posto pari al valore medio, riducendo la varianza complessiva 10/. Analoghe considerazioni valgono per lo stimatore (4). Pertanto, al fine di limitare la distorsione nella stima della variabilità delle informazioni è preferibile l'uso di metodi di imputazione casuale del tipo (5) 11/.

Per quanto riguarda la tecnica (6) si osservi che la distorsione nella stima della variabilità dipende dalla capacità esplicativa della funzione g(.)

utilizzata per l'imputazione; in particolare la distorsione risulta nulla quando tale funzione prevede esattamente, date le variabili esplicative, il valore della variabile non indicata.

Per il modello 2, così come accade per la stima del valore medio, la variabilità stimata risulta generalmente inferiore al valore calcolato per la popolazione di riferimento; questo fenomeno si verifica per tutte le tecniche ad eccezione della (1); in tale modello la probabilità di non risposta (direttamente correlata con il valore dell'immobile) riduce la variabilità dei dati relativi ai rispondenti in quanto le informazioni perdute sono quelle che si posizionano all'estremità destra della distribuzione.

3.1 I metodi di imputazione nell'indagine sugli investimenti delle imprese industriali

L'indagine sugli investimenti delle imprese industriali è condotta su un campione - stratificato per classe dimensionale (in termini di addetti) e ramo di attività economica - di circa 1.000 aziende produttive del settore manifatturiero. Il campione è costruito in base alla distribuzione bivariata (per dimensione e settore di attività) degli impianti industriali appartenenti a imprese con 20 addetti ed oltre; la quota di impianti oggetto di rilevazione, rispetto a quelli dell'universo di riferimento, è variabile tra gli strati e aumenta al crescere della dimensione dell'impresa (Banca d'Italia, 1987b).

L'indagine in esame è finalizzata prevalentemente allo studio della dinamica dell'investimento e dell'occupazione e, solo secondariamente, alla stima del livello di queste variabili. Risulta pertanto evidente che le usuali metodologie di imputazione delle risposte mancanti (ad esempio hot-deck) determinano una distorsione nella stima della covarianza tra i valori delle variabili ai tempi t_0 e t_1 e quindi una stima non completamente affidabile della variazione intervenuta nel periodo 12/.

Escludendo a priori sia i modelli a nonignorable response mechanism (Little, 1983, pg. 383) - che presuppongono la conoscenza del legame tra il valore della variabile da imputare e la probabilità di non risposta - sia i modelli che richiedono la normalità della distribuzione multivariata delle variabili da imputare 13/, la scelta delle tecniche da sperimentare risulta limitata alle seguenti possibilità:

(1) hot-deck per l'imputazione dei tassi di variazione: la variabile al tempo t_1 viene stimata mediante il prodotto tra la stessa, rilevata al tempo t_0 e il tasso di variazione, imputato con una estrazione casuale nella stessa cella di imputazione ;

(2) metodo dell'impresa tipo: le informazioni mancanti sono integrate con i dati relativi ad una impresa operante nello stesso settore di attività e appartenente alla stessa classe dimensionale;

(3) metodo del tasso di variazione medio (la descrizione del metodo è riportata nell'appendice "B"): l'imputazione dei tassi di variazione viene effettuata sulla base del tasso di variazione medio delle imprese operanti nello stesso settore di attività e appartenenti alla stessa classe dimensionale ed alla stessa area geografica; la ricostruzione della variabile mancante avviene come per il metodo (1);

(4) post-stratificazione (la descrizione del metodo è riportata nell'appendice "B"): l'insieme delle variabili rilevate viene ripartito in sottoinsiemi escludendo i MD dall'analisi. Le imprese restanti sono riponderate per classe dimensionale e settore di attività economica in modo da ricostruire la stessa distribuzione dell'universo di riferimento.

Da recenti analisi condotte per l'indagine annuale ISTAT sul prodotto lordo (Quintano et Alii, 1986), risulta che il metodo (2) è qualitativamente inferiore al metodo (3); inoltre, poiché l'indagine condotta dalla Banca ha come obiettivo la stima delle variazioni medie delle variabili oggetto di indagine e

non l'analisi della funzione di distribuzione dei tassi di variazione, il metodo (1) risulta meno idoneo dei metodi (3) e (4) in quanto fornisce, in media e nell'ipotesi che ciascuna unità campionaria sia caratterizzata dalla stessa probabilità di non risposta, gli stessi risultati di questi ultimi, ma causa un aumento della variabilità delle informazioni 14/.

Le procedure sperimentate e poste a confronto sono pertanto la (3) e la (4).

3.2 La valutazione delle procedure

Allo scopo di valutare quale procedura risulta più efficiente per la stima del livello e della dinamica dei valori medi delle variabili relative all'investimento (descritte nell'appendice "C"), rilevate con il campione 1986, si è estratto il sottoinsieme di unità statistiche che hanno fornito informazioni su tutti i fenomeni oggetto di analisi. Tale sottoinsieme, assunto come popolazione di riferimento, ha consentito di determinare i valori medi "veri" del livello e della dinamica delle variabili in esame.

In analogia con l'analisi svolta per il patrimonio in immobili, sono state esaminate le caratteristiche dei MD, stimando le probabilità di non risposta relative a ciascuna variabile sulla base del rapporto tra il numero di MD presenti nel campione e la numerosità campionaria 15/.

Inoltre, poiché il metodo (3) è principalmente fondato sulla possibilità di disporre di informazioni per almeno uno dei due periodi oggetto di rilevazione, sono state stimate, per coppie di variabili, le probabilità congiunte di non risposta. Ciò ha consentito di generare casualmente MD con una struttura simile a quella osservata per il complesso del campio-

ne; la successiva applicazione delle procedure di stima (o imputazione) (3) e (4) e l'iterazione del procedimento hanno infine permesso di stimare, per ogni variabile, le distorsioni determinate dai due metodi 16/.

Dai risultati riportati nella tavola 3 si osserva che le stime ottenute con il metodo (3) risultano caratterizzate da distorsioni percentuali inferiori (in valore assoluto) ai corrispondenti valori ottenuti con il metodo (4), sia per il livello che per la dinamica dell'investimento. Per quanto riguarda l'errore quadratico medio relativo, si nota che, fatta eccezione per la stima della dinamica dell'investimento previsto per il 1987, l'errore dello stimatore (3) è inferiore al corrispondente valore ottenuto in base alla procedura (4). Il metodo del tasso di variazione medio risulta pertanto preferibile al metodo di ponderazione.

3.3 La violazione delle ipotesi

La sperimentazione delle procedure (3) e (4), riportata nel precedente paragrafo, è stata condotta nell'ipotesi che, all'interno delle celle di imputazione, ogni unità statistica fosse caratterizzata dalla medesima probabilità di non risposta. In realtà non si dispone di informazioni per valutare in che misura tale ipotesi sia fondata. Si è deciso pertanto di sperimentare il comportamento dei due stimatori nel caso di violazione delle ipotesi, incrementando arbitrariamente, rispetto alla precedente analisi, le probabilità di non risposta di alcune unità. In particolare, per le imprese di minore dimensione appartenenti alle classi "200-499 addetti" e "500-999 addetti", il numero di MD è stato aumentato del 100 per

Tav.3

Confronto tra il metodo di riponderazione ed il metodo del tasso di variazione
(probabilità di non risposta costante ($))

Variabile	Distorsione percentuale(*) metodo(3)	metodo (4)	SQD(**) metodo(3)	metodo(4)
v200	+ 0,1	- 1,1	0,2	10,7
v201	- 0,1	- 4,8	0,4	12,3
v202	- 0,2	+ 0,4	0,5	10,2
v203	- 1,9	- 2,9	4,3	10,6
v205	- 0,0	- 2,0	0,1	6,5
DO86	+ 0,0	+ 2,5	0,1	7,0
DIE86	- 0,3	+ 1,5	0,5	2,1
DIP86	- 0,2	- 3,9	0,6	5,6
DIP87	- 1,7	- 3,3	4,4	3,7

($)All'interno delle celle di imputazione.
(*)Differenza tra il valore medio stimato (m) ed il valore medio "vero" (M), in percentuale del valore medio "vero".
(**)$SQD=(\hat{E}(m-M)^2/M^2)^{0.5} \times 100$

Tav.4

Confronto tra il metodo di riponderazione ed il metodo del tasso di variazione
(probabilità di non risposta variabile($))

Variabile	Distorsione percentuale(*) metodo(3)	metodo (4)	SQD(**) metodo(3)	metodo(4)
v200	- 3,3	+ 3,0	6,9	15,8
v201	- 1,9	- 0,9	4,9	13,6
v202	- 1,1	+ 4,3	2,6	16,6
v203	- 2,3	+ 0,1	4,3	11,5
v205	- 0,0	+ 0,6	0,1	6,8
DO86	+ 0,0	- 0,2	0,1	6,8
DIE86	+ 2,8	+ 1,3	8,3	3,3
DIP86	+ 1,7	- 3,4	5,9	7,3
DIP87	- 1,2	- 3,4	3,2	7,7

($)All'interno delle celle di imputazione.
(*)Differenza tra il valore medio stimato (m) ed il valore medio "vero" (M), in percentuale del valore medio "vero".

$$(**)SQD=(\hat{E}(m-M)^2/M^2)^{0.5} \times 100$$

cento circa 17/.

I risultati di questa seconda elaborazione sono riportati nella tavola 4. Si osserva che sia lo stimatore (3) che lo stimatore (4) risentono della violazione delle ipotesi; pertanto, gli errori quadratici risultano superiori rispetto al caso precedente. Il comportamento dei due stimatori si presenta tuttavia notevolmente differenziato; il metodo (3) tende infatti a sottostimare i valori medi dell'investimento, a differenza del metodo (4), che presenta un comportamento tendenzialmente opposto. Nell'esperimento condotto, la probabilità di non risposta delle imprese di minore dimensione, all'interno di ciascuna cella di imputazione, è infatti superiore alla probabilità media di non risposta, relativa alla stessa cella. La correlazione positiva tra il livello dell'investimento e la dimensione dell'impresa (in termini di addetti), ha pertanto contribuito alla sovrastima dei valori medi delle variabili imputate con il metodo della riponderazione, rispetto a quelli calcolati con il metodo del tasso di variazione.

4. Conclusioni

Per quanto riguarda il patrimonio in immobili, come evidenziato nei precedenti paragrafi, il metodo 2.1.(5), basato sull'imputazione casuale, fornisce i migliori risultati, sia per quanto riguarda la stima del valore medio che per la valutazione della variabilità delle informazioni con le usuali procedure di calcolo. Inoltre, tale metodo può essere generalizzato, per tenere conto della distorsione che l'imputazione casuale delle informazioni mancanti determina nella stima di covarianze e correlazioni.

Per quanto attiene la gestione operativa di una indagine campionaria, è evidente il legame esistente tra la qualità dell'informazione e la complessità del metodo di imputazione; in proposito si nota che, nel caso dell'indagine sui bilanci delle famiglie, la maggior parte della distorsione determinata dalla presenza di MD risulta eliminata utilizzando le tecniche più semplici. La modesta percentuale di MD presenti nelle componenti elementari della ricchezza reale e, soprattutto, le modeste differenze tra le probabilità di non risposta per le tipologie di immobili considerate rendono infatti apprezzabili i risultati ottenuti sostituendo ai valori mancanti il valore medio calcolato per i rispondenti (2.1.(3) e 2.1.(4)). Tali stime risultano comunque qualitativamente inferiori rispetto a quelle che si ottengono utilizzando metodi di imputazione casuale (2.1.(5)) o metodi di regressione (2.1.(6)).

Per quanto riguarda l'indagine sugli investimenti, la procedura basata sull'imputazione del tasso di variazione medio (3.1.(3)) risulta preferibile al metodo di riponderazione (3.1.(4)), in quanto consente di ottenere stime caratterizzate da minori distorsioni e maggiore affidabilità. Inoltre, a differenza del metodo 3.1.(3), la procedura di riponderazione è piuttosto sensibile all'incremento della probabilità media di non risposta, in quanto sono escluse dall'analisi tutte le imprese con almeno un MD; ciò può determinare una rilevante riduzione della numerosità campionaria.

Infine, è opportuno evidenziare i limiti delle procedure di imputazione esaminate. Poiché l'obiettivo di queste consiste nella stima di valori medi, i singoli valori imputati non possono essere utilizzati come stime puntuali dei valori mancanti; a questo fine sarebbe interessante sperimentare metodi di imputazione basati sulla regressione multivariata (Little, 1983, pg.367).

Note

1/ I questionari mancanti delle componenti elementari della variabile reddito sono restituiti agli intervistatori per il completamento, da effettuarsi contattando nuovamente la famiglia. Nel caso che l'intervistato rifiuti di fornire informazioni su questa variabile, il questionario relativo all'intera famiglia viene escluso da ogni elaborazione.

2/ La ricchezza in immobili costituisce circa il 90 per cento della ricchezza reale complessiva.

3/ Una analisi simile è stata condotta per la National Farm Survey in Canada (Cheung and Seko,1986), utilizzando come popolazione di riferimento il campione complessivo, contenente valori imputati per mezzo di una procedura hot-deck sequenziale. Nel presente lavoro si è preferito eliminare i missing data, per evitare distorsioni determinate dalle procedure di imputazione nella stima della covarianza tra il valore dell'immobile, il tipo di quest'ultimo e le caratteristiche familiari (Kalton and Kasprzyk, 1986).

4/ Sulla base di tale modello la probabilità **p** di generare M missing data su n osservazioni appartenenti a uno dei gruppi selezionati in relazione a un vettore di parametri **y** risulta la seguente:

$$p(M \mid \mathbf{y}) = \binom{n}{M} P(\mathbf{y})^{M} (1-P(\mathbf{y}))^{(n-M)}$$

dove:

$$P(\mathbf{y}) = \frac{1}{1 + e^{-(a_0 + \mathbf{a}'\mathbf{y})}}$$

5/ Formalmente, la probabilità di non risposta risulta pari a:

$$P = \frac{1}{1 + e^{-(A+B*eta+C*educ+D*tipo1+E*tipo2)}}$$

dove:

eta =età del capofamiglia
educ =livello di istruzione del CF (anni di studio)

tipo1 =dummy (0,1) che identifica l'abitazione dove vive la famiglia
tipo2 =dummy (0,1) che identifica le altre abitazioni e i fabbricati di proprietà della famiglia.

I parametri del modello, stimati con il metodo dei minimi quadrati non lineari, risultano i seguenti

Parametro	valore stimato	Intervallo di confidenza (p=0,95)	
A	- 2,85	- 3,406	- 2,293
B	0,02	0,009	0,035
C	- 0,04	- 0,069	- 0,012
D	- 0,76	- 0,980	- 0,538
E	- 0,75	- 1,066	- 0,444

6/Negli esperimenti condotti :

$$g(Y\ P)= \alpha+\beta_1 y_1+\beta_2 y_2+\beta_3 y_3+\beta_4 y_4+\gamma P$$

Le variabili P, y_1=tipo1, y_2=tipo2, y_3=eta e y_4=educ sono descritte nella nota 5/.

7/In generale, le differenze tra rispondenti e non rispondenti determinano distorsioni nelle stime dei valori medi (Madow, Nisselson, Olkin (1983, vol. I, pg. 25); tuttavia, se la probabilità di non risposta non è correlata con le variabili rilevate, è possibile assumere che la media calcolata sui rispondenti sia uguale alla media relativa ai non rispondenti (Ford, 1983).

8/ Nel caso di due sottoinsiemi caratterizzati da una probabilità di non risposta rispettivamente pari a zero e a P il valore medio dello stimatore risulta:

$$E(\bar{x}_2)=[1/(N-N_2P\ )][N_1E_1+N_2(1-P\ )E_2]$$

dove E_i (i=1,2) rappresenta la media relativa ai due sottoinsiemi di numerosità N_i (i=1,2), e la distorsione è pari a:

$$E(\bar{x}_2)-E(x)= \frac{N_1\ N_2\ P\ (E_1- E_2)}{(N-N_2P)N}$$

Tale distorsione è maggiore di zero quando la media relativa al secondo sottoinsieme è inferiore alla media relativa al primo.

9/Si fa riferimento alla variabilità delle informazioni, in quanto tale valore è necessario per la stima della varianza della media campionaria. Le informazioni mancanti (imputate) vengono trattate, nella presente analisi, nello stesso modo di quelle rilevate; ciò al fine di valutare la distorsione determinata dalle usuali procedure di calcolo della varianza in presenza di MD imputati.
La presenza di non risposte ha comunque ulteriori effetti, oltre alla distorsione della stima della varianza delle informazioni, sulla varianza della media campionaria; tali effetti derivano sia dall'effettiva riduzione del campione che dalle tecniche utilizzate per il trattamento dei MD (Hansen, Hurwitz, Madow, 1953; Lock Oh, Scheuren, 1983; Ford, 1983).

10/In questo caso l'ampiezza dell'intervallo di confidenza del valore medio risulta distorta del fattore R/N (Madow, Nisselson, Olkin, 1983 pg. 89).

11/Utilizzando le informazioni imputate per la stima dell'intervallo di confidenza si determina una distorsione della stima dell'ampiezza dell'intervallo pari a $[2(n-m)/n]x100$ %, dove n rappresenta l'ampiezza del campione ed m il numero di rispondenti (Ford, 1983, pg. 192).

12/Come evidenziato da Santos(1981), la distorsione relativa della stima della covarianza tra due variabili x e y è approssimativamente pari al tasso di non risposta, quando si utilizzano tecniche di imputazione casuale. In questo caso infatti i valori imputati di y sono incorrelati con i corrispondenti valori di x (e viceversa); ciò determina una riduzione (in valore assoluto) della covarianza.

13/La distribuzione delle imprese per classi di investimento è asimmetrica e decresce lentamente (heavy-tailed) all'aumentare del livello dell'investimento.

14/Utilizzando il metodo (3) tutte le imprese che non hanno fornito informazioni sul valore degli investimenti, in uno qualsiasi dei due periodi (t_0 e t_1) oggetto di rilevazione, risultano caratterizzate da un tasso di variazione (imputato) pari alla media di tassi di variazione calcolati per le restanti unità campionarie. La distribuzione delle imprese per tassi di variazione degli investimenti risulta quindi sovrastimata in corrispondenza del tasso di variazione medio.

15/Per le variabili esaminate, il tasso di non risposta

stimato sui dati campionari risulta compreso nell'intervallo 2,5 - 4,0 per cento.

16/Il procedimento di generazione di MD e di successiva imputazione è stato replicato 10 volte, di cui 5 nell'ipotesi che le probabilità di non risposta fossero costanti all'interno delle celle di imputazione e 5 ipotizzando probabilità di non risposta variabili (cfr. paragrafo 3.3).

17/Indicata con P_{ij} la probabilità di non risposta relativa alla j-esima variabile osservata su un'impresa appartenente alla i-esima classe dimensionale, la procedura di generazione casuale dei MD si modifica, rispetto all'analisi riportata nel paragrafo 3.2, nel modo seguente:

Classe dimensionale (addetti)	Sottoclasse	Probabilità di non risposta	
		Ipotesi par.3.2	Ipotesi par.3.3
200 - 499	200 - 299	P_{3j}	$2xP_{3j}$
	300 - 499	P_{3j}	P_{3j}
500 - 999	500 - 749	P_{4j}	$2xP_{4j}$
	750 - 999	P_{4j}	P_{4j}

Operando in tal modo, la probabilità media di non risposta, calcolata per il complesso del campione, incrementa del 35 per cento circa.

APPENDICE "A"

Tabella dei simboli utilizzati

Σ	Simbolo di sommatoria
N	Numero complessivo di osservazioni
R	Numero complessivo di risposte
L	Numero di celle di imputazione
N_h	Numero di osservazioni nella cella h
R_h	Numero di risposte nella cella h
N_h-R_h	Numero di missing data nella cella h
x_{Ri}	Valore della variabile x per l'i-esimo rispondente
x_{Rhi}	Valore della variabile x per l'i-esimo rispondente della cella h
P_h	Probabilità di non risposta nella cella h-esima
$\overline{P}$	Probabilità media di non risposta
x_{NRj}	Valore della variabile x per il j-esimo non rispondente
$x"_{NRj}$	Stima di x_{NRj} ottenuta in funzione delle variabili disponibili
x_i	Valore della variabile x per la i-esima unità di rilevazione (il valore non è noto se l'unità statistica rifiuta di rispondere).
x_{hi}	Valore della variabile x per la i-esima unità di rilevazione nella h-esima cella di imputazione

APPENDICE "B"

Il metodo di post-stratificazione

In presenza di MD, la stima della covarianza tra due variabili x e y o del tasso di variazione di un fenomeno (ad esempio x_t/x_{t-1}) può essere effettuata trascurando le unità campionarie che presentano un MD in uno qualsiasi dei due campi oggetto di analisi (ad esempio x o y) e riponderando le restanti unità in modo da ricostruire la distribuzione (per uno o più caratteri rilevanti) dell'universo di riferimento.

All'aumentare della numerosità dell'insieme di variabili oggetto di rilevazione, il procedimento sopra descritto incrementa sensibilmente i tempi di elaborazione, in quanto è necessario operare una selezione e una riponderazione per il calcolo di ogni covarianza. Al contrario, l'eliminazione delle unità campionarie che presentano un MD in un campo qualsiasi riduce sensibilmente i tempi di elaborazione (essendo richiesta un'unica selezione e un'unica riponderazione) ma può determinare una inaccettabile riduzione della numerosità campionaria.

Il metodo utilizzato rappresenta un compromesso tra la dimensione del campione e la riduzione dei tempi di elaborazione. In particolare, sono stati definiti due sottoinsiemi, contenenti rispettivamente le variabili v205 e v206 e le variabili v200, v201, v202, v203. Le imprese che hanno fornito informazioni su tutte le variabili del sottoinsieme sono state riponderate in modo da ricostruire la distribuzione (per classe dimensionale e settore di attività economica) dell'universo di riferimento.

Il metodo del tasso di variazione medio

Il metodo si basa sull'aspetto longitudinale dell'indagine; ad ogni impresa sono richieste informazioni relative all'esercizio corrente ed a quello passato. Pertanto, ripartendo le imprese in gruppi distinti per classe dimensionale, settore di attività economica ed area geografica, è possibile calcolare il valore medio per addetto di ogni variabile rilevata, per ciascuno dei due esercizi:

$$PCM(t) = \Sigma_i \; PC(i,t)\, ADD(i,t) / \Sigma_i \; ADD(i,t)$$

dove PCM(t), PC(i,t), ADD(i,t) rappresentano rispettivamente il valore medio procapite al tempo t, il valore procapite per l'impresa i al tempo t ed il numero di addetti dell'impresa i al tempo t. La stima del tasso di variazione del valore procapite medio risulta pari a:

$$VAR = [PCM(t)-PCM(t-1)]/PCM(t-1)$$

di conseguenza i valori mancanti (VM(i,t)) possono essere ricostruiti nel modo seguente:

$$VM(i,t)= PC(i,t-1)x(1+VAR)xADD(i,t)$$

Per le informazioni che non possono essere imputate con il metodo sopra descritto (ad esempio quando non si dispone di nessuno dei due valori, ai tempi t_0 e t_1, della variabile da imputare) si è fatto ricorso al valore medio calcolato per le imprese appartenenti alla stessa cella di imputazione.

La descrizione dettagliata del piano di ricostruzione dei valori mancanti è riportata nell'appendice "C".

APPENDICE "C"

Elenco delle variabili

v200	spesa per investimenti sostenuta nel 1985
v201	spesa per investimenti prevista a fine '85 per il 1986
v202	spesa per investimenti sostenuta nel 1986
v203	spesa per investimenti prevista per il 1987
v205	occupazione a fine '85
v206	occupazione a fine '86

DO86	v206/v205
DIE86	v202/v200
DIP86	v201/v200
DIP87	v203/v202

Metodi di imputazione (*) (**)

Variabile	Metodo
v206	la variabile non assume valori missing
v205	v205=v206xM(v205)/M(v206)
v202	se v200 è disponibile e diversa da zero: v202=[PC(v200)xPCM(v202)/PCM(v200)]xv206
v200	se v202 è disponibile e diversa da zero: v200=[PC(v202)xPCM(v200)/PCM(v202)]xv205
v202	se v200=0: v202=v206xPCM(v202)
v200	se v202=0: v200=v205xPCM(v200)
v202 v200	se v201 è disponibile e diversa da zero: v202=[PC(v201)xPCM(v202)/PCM(v201)]xv206 v200=[v202/v206xPCM(v200)/PCM(v202)]xv205 se v201 non è disponibile o uguale a zero: v202=v206xPCM(v202) v200=[v202/v206xPCM(v200)/PCM(v202)]xv205
v203	se v202 è disponibile e diversa da zero: v203=[v202/v206xPCM(v203)/PCM(v202)]xv206
v201	se v200 è disponibile e diversa da zero: v201=PC(v200)x[PCM(v201)/PCM(v200)]xv205
v203	se v202=0: v203=v206xPCM(v203)
v201	se v200=0: v201=v205xPCM(v201)

(*)Le celle di imputazione sono determinate dall'intersezione delle modalità:: classe dimensionale, settore di attività economica, area geografica.

(**)I simboli **M(x)**, **PCM(x)** rappresentano rispettivamente il valore medio della variabile x ed il valore medio della variabile x rapportato al numero medio di addetti. Il livello di disaggregazione per il calcolo di tali valori medi è rappresentato

dall'intersezione delle modalità: classe dimensionale, settore di attività economica, area geografica. Il simbolo **PC(x)** rappresenta il rapporto tra la variabile x e e il numero di addetti dell'impresa.

RIFERIMENTI BIBLIOGRAFICI

BANCA D'ITALIA (1987a) I bilanci delle famiglie italiane nell'anno 1986 (a cura di Cannari L.), "Bollettino Statistico" n.1-2, gennaio-giugno 1987, Roma.

BANCA D'ITALIA (1987b) Assemblea Generale Ordinaria dei Partecipanti, Appendice, pg. 233, Tav. B 11., Roma.

CHEUNG S. and SEKO C. (1986) A Study of the Effects of Imputation Groups in the Nearest Neighbour Imputation Method of the National Farm Survey, in "Survey Methodology" 12, 1, Statistics Canada.

DAVID M.H., LITTLE R., SAMUHEL M. and TRIEST R. (1983) Imputation Models Based on the Propensity to responde, in "Proceedings of the Business and Economic Statistics Section", A.S.A..

FORD B.L. (1983) An Overview of hot-deck procedures in "Incomplete Data in Sample Survey", vol. II, pg. 191, Academic Press, New York.

HUNSEN M.H., HURWITZ W.N., MADOW W.G. (1953) "Sample Survey Methods and Theory", Wiley, New York.

KALTON G. and KASPRZYK D. (1986) The Treatment of Missing Survey Data, in "Survey Methodology", 12, 1, Statistics Canada.

LITTLE R.J.A. (1983) The Nonignorable Case in "Incomplete Data in Sample Survey", vol. II, pg. 389, Academic Press, New York.

LOCK OH H. SCHEUREN F.J.(1983) Weighting Adjustment for Unit Nonresponse in "Incomplete Data in Sample Survey", vol. II, pg. 146, Academic Press, New York.

MADOW W.G. NISSELSON H. OLKIN I. (1983) "Incomplete Data in Sample Survey", vol. I, Academic Press New York.

PLATEK R. and GRAY G.B. (1983) Imputation Methodology in "Incomplete Data in Sample Survey", vol. II, pg. 283, Academic Press, New York.

QUINTANO C. CALZARONI M. DINI P. MASSELLI M. POLITI M. TACCINI P. (1986) "Una ricognizione dell'error profile dell'indagine ISTAT sul prodotto lordo", dattiloscritto.

RUBIN D.B. (1983) Conceptual Issues in the Presence of Nonresponse in "Incomplete Data in Sample Survey", vol. II, pg. 123, Academic Press, New York.

SANTOS R.L. (1981) Effects of Imputation on Regression Coefficients, in "Proceedings of the Section on Survey Research Methods", A.S.A..

ELENCO DEI PIÙ RECENTI TEMI DI DISCUSSIONE (*)

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. CORAGGIO - 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 e T. RAGANELLI (febbraio 1988).

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

BANCA D'ITALIA - CENTRO STAMPA