

BANCA D'ITALIA

Temi di discussione

del Servizio Studi

L'indagine sui bilanci delle famiglie italiane nel 1984

di Luigi Cannari

Numero 86 - Aprile 1987

BANCA D'ITALIA

Temi di discussione

del Servizio Studi

L'indagine sui bilanci delle famiglie italiane nel 1984

Alcune considerazioni sugli errori campionari

di Luigi Cannari

Numero 86 - Aprile 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, FRANCO COTULA, IGNAZIO VISCO, STEFANO VONA;
MARIA ANTONIETTA ORIO (segretaria).*

SOMMARIO

Nel presente lavoro si valutano le caratteristiche di taluni stimatori del totale delle principali grandezze rilevate con l'indagine sui bilanci delle famiglie italiane, condotta dalla Banca d'Italia con riferimento all'anno 1984.

L'efficienza degli stimatori utilizzati e l'opportunità di ricorrere a procedure di stratificazione a posteriori è valutata mediante la stima degli errori campionari, effettuata per disaggregazione territoriale e per il totale della popolazione.

Il calcolo dei suddetti errori, condotto sulla base del piano di campionamento utilizzato (a due stadi con stratificazione delle unità di primo stadio) consente infine la valutazione dell'effetto complessivo del disegno campionario e delle differenze che intercorrono, nel campione considerato, tra le stime degli errori standard nell'ipotesi di campionamento casuale semplice e le stesse riferite a un disegno campionario complesso.

INDICE

Introduzione	pag. 5
1. La variabilità dell'ampiezza delle unità di primo stadio	pag. 6
2. La stratificazione a posteriori	pag. 13
3. Conclusioni	pag. 19
Appendice	pag. 23
Bibliografia	pag. 25

Introduzione(*)

Con l'indagine sui bilanci delle famiglie italiane nel 1984 la Banca d'Italia ha acquisito le informazioni necessarie per la valutazione degli errori campionari delle grandezze rilevate; in particolare sono state codificate e memorizzate nella base dati, a differenza di quanto avveniva nelle indagini precedenti, informazioni sul comune di appartenenza delle famiglie intervistate e sulla probabilità di estrazione di ciascuna unità di campionamento.

Poichè la stima degli errori campionari, come noto, è di notevole importanza ai fini dell'analisi statistica dei risultati di una ricerca, nel presente lavoro si è valutato, con riferimento a diversi stimatori e per disaggregazione territoriale, la variabilità delle principali grandezze oggetto di indagine.

Il lavoro è costituito da tre paragrafi: nel primo vengono stimati gli effetti che la variabilità della dimensione delle unità primarie di campionamento determina sulla varianza delle stime dei totali; nel secondo viene valutata l'opportunità di ricorrere a procedure di stratificazione a posteriori; nel terzo, infine, sono riportate le principali conclusioni.

Per agevolare il lettore nella comprensione delle formule riportate nel testo, è stata inserita in appendice una tabella dei simboli più frequentemente utilizzati.

(*) L'autore desidera ringraziare Luigi Biggeri, Carlo Andrea Bollino, Lorenzo Fattorini, Daniela Gressani, Achille Lemmi e Rocco A. Pirròtta per i numerosi commenti ricevuti.

1. La variabilità dell'ampiezza delle unità di primo stadio

Nel campionamento a due stadi, stratificato al primo stadio con le unità primarie (UP) selezionate con probabilità costanti e senza reimmissione e con unità secondarie (US) estratte in ogni UP con uguale probabilità, una stima corretta del totale T 1/ della popolazione è data dall'espressione:

$$(1) \quad \hat{T} = \sum_h^L M_h / m_h \sum_1^{m_h} N_{hi} \bar{x}_{hi}$$

dove:

$$\bar{x}_{hi} = 1/n_{hi} \sum_j^{n_{hi}} x_{hij}$$

la cui varianza, nel caso in cui le US siano estratte senza reimmissione, è definita dalla seguente relazione :

$$(2) \quad V(\hat{T}) = \sum_h^L M_h^2 / m_h (M_h - m_h) / M_h S_{1h}^2 + \sum_h^L M_h / m_h \sum_1^{m_h} N_{hi}^2 / n_{hi} \cdot (N_{hi} - n_{hi}) / N_{hi} S_{2hi}^2$$

dove:

$$S_{1h}^2 = 1/(M_h - 1) \sum_1^{M_h} (T_{hi} - T_h / M_h)^2$$

$$S_{2hi}^2 = 1/(N_{hi} - 1) \sum_j^{N_{hi}} (x_{hij} - T_{hi} / N_{hi})^2$$

Quando il totale di una determinata caratteristica della popolazione viene calcolato come il prodotto tra la

stima del totale medio per unità primaria di campionamento e il numero complessivo di UP, ovvero con la formula (1), la variabilità della dimensione delle UP può avere effetti sensibili sulla varianza delle stime 2/; infatti, tanto più il numero complessivo di US (N_{hi}) per UP è soggetto a una elevata variabilità nell'universo delle M ($M = \sum_h^L M_h$) unità primarie di campionamento, tanto più è probabile estrarre un campione che generi una stima della numerosità media di US, per le UP estratte, lontana dal valore vero (cioè $\sum_i^{M_h} N_{hi} / M_h$ può essere notevolmente diversa da $\sum_i^{m_h} N_{hi} / m_h$). Poiché detta numerosità influenza direttamente la stima del totale medio per le UP estratte ($1/m_h \sum_i^{m_h} N_{hi} \bar{x}_{hi}$) anche quest'ultima stima risulta caratterizzata da una notevole variabilità; ciò determina una perdita di efficienza per gli stimatori del tipo (1).

Al fine di valutare la rilevanza di questo problema per quanto riguarda il campione B.I. per l'indagine sui bilanci delle famiglie italiane nell'anno 1984, è stata condotta una analisi preliminare sull'ampiezza delle UP campionate. L'analisi ha evidenziato che il numero medio di famiglie per UP (a livello di strato) risultante dai dati campionari è generalmente superiore al corrispondente valore calcolato per il complesso dei comuni italiani (a livello di strato). Questo fenomeno si verifica nonostante i criteri di stratificazione adottati tengano conto dell'ampiezza delle UP; il disegno campionario utilizzato prevede infatti la ripartizione del territorio nazionale in 85 strati (regioni per classi di ampiezza demografica dei comuni). L'estrazione delle unità primarie all'interno degli strati viene effettuata con campionamento casuale, assegnando a ciascun comune la stessa probabilità di estrazione.

Questo metodo non risulta efficiente per il controllo degli effetti della variabilità della dimensione delle UP 3/, per il quale sarebbe preferibile l'uso di

altri metodi, come ad esempio l'estrazione delle UP con probabilità proporzionali all'ampiezza demografica dei comuni 4/. Inoltre il criterio di selezione utilizzato contribuisce alla sovrarappresentazione, nel campione considerato, di UP di dimensioni maggiori, in termini di numerosità di US; infatti le unità primarie, in origine estratte casualmente, sono state in parte successivamente sostituite per esigenze legate alla difficoltà di disporre di una rete di rilevazione efficiente in alcuni comuni, frequentemente i più piccoli.

Dato il piano di campionamento, per migliorare l'efficienza degli stimatori del totale si può ricorrere a stimatori meno sensibili alla variabilità della dimensione delle UP, come lo stimatore ottenuto dalla (1) riponderandola con il rapporto tra N e $\hat{N}$; tale stimatore infatti non risente particolarmente della varianza della dimensione delle UP poiché sia $\hat{N}$ che $\hat{T}$ sono caratterizzati da una analoga distorsione. Formalmente sia:

$$(3) \quad {}_1\hat{T} = N/\hat{N} \hat{T}$$

dove $\hat{T}$ è dato dalla (1) e $\hat{N}$, la stima della numerosità della popolazione, è pari a:

$$\hat{N} = \sum_h^L M_h/m_h \sum_i^{m_h} N_{hi}$$

Si noti che ${}_1\hat{T}$ è uno stimatore distorto, ma asintoticamente corretto di T in quanto il bias, pari a $(E(\hat{T}/\hat{N}) - E(\hat{T})/E(\hat{N}))$, converge in probabilità a zero 5/.

Il confronto di ${}_1\hat{T}$ con $\hat{T}$ - allo scopo di valutarne l'efficienza relativa - è stato effettuato sulla base del momento secondo dell'errore dello stimatore (in seguito MSE), che per $\hat{T}$ equivale alla (2) e per ${}_1\hat{T}$ è pari a:

$$(4) \quad \text{MSE}(\hat{T}) = E(\hat{T} - N E(\hat{T})/E(N))^2 =$$

$$= \sum_h^L M_h^2/m_h(M_h-m_h)/M_h S_{1h}^2 + \sum_h^L M_h/m_h \sum_i^{N_h} N_{hi}^2/n_{hi}(N_{hi}-n_{hi})/N_{hi} S_{2hi}^2$$

dove:

$$S_{1h}^2 = 1/(M_h-1) \cdot \sum_1^{M_h} (T_{hi} - T_h/M_h)^2 + (T/N)^2 \cdot 1/(M_h-1) \sum_1^{M_h} (N_{hi} - N_h/M_h)^2 +$$

$$- 2 T/N \cdot 1/(M_h-1) \sum_1^{M_h} (T_{hi} - T_h/M_h)(N_{hi} - N_h/M_h)$$

$$S_{2hi}^2 = 1/(N_{hi}-1) \sum_j^{N_{hi}} (x_{hij} - T_{hi}/N_{hi})^2$$

La differenza (D_1) tra la varianza di $\hat{T}$ e il MSE di $\hat{T}$ risulta pertanto pari a :

$$(5) \quad D_1 = V(\hat{T}) - \text{MSE}(\hat{T}) = \sum_h^L M_h^2/m_h(M_h-m_h)/M_h \cdot 1/(M_h-1) \cdot T/N \cdot$$

$$\cdot (-T/N \sum_1^{M_h} (N_{hi} - N_h/M_h)^2 + 2 \sum_1^{M_h} (T_{hi} - T_h/M_h)(N_{hi} - N_h/M_h))$$

e può essere stimata con la formula seguente 6/:

$$(6) \quad \hat{D}_1 = \hat{V}(\hat{T}) - \hat{\text{MSE}}(\hat{T}) = \sum_h^L M_h^2/m_h(M_h-m_h)/M_h \cdot 1/(m_h-1) \cdot \hat{T}/\hat{N} \cdot$$

$$\cdot (-\hat{T}/\hat{N} \sum_1^{m_h} (N_{hi} - \hat{N}_h/M_h)^2 + 2 \sum_1^{m_h} (\hat{T}_{hi} - \hat{T}_h/M_h)(N_{hi} - \hat{N}_h/M_h))$$

La stima della differenza del MSE determinata dalla applicazione dello stimatore $\hat{T}_1$ proposto nella (3) in luogo di $\hat{T}$ proposto nella (1) è stata quindi effettuata sui dati campionari B.I. del 1984. Gli strati originari 7/ sono stati raggruppati suddividendo l'intero territorio nazionale in 12 ripartizioni geografiche e mantenendo inalterato il parametro di stratificazione per "classe di ampiezza demografica dei comuni"; il numero di strati considerati si è pertanto ridotto da 85 a 60 (12 ripartizioni per 5 classi di ampiezza). Ciò ha determinato una sovrastima della varianza dei totali 8/, per il minor peso dell'effetto stratificazione; questo procedimento si è reso peraltro necessario per tutti gli strati contraddistinti da una sola UP campionata (per i quali non era possibile calcolare la varianza dei totali comunali) e opportuno per gli strati con due sole UP estratte; in quest'ultimo caso infatti la correlazione stimata tra $\hat{T}_{h_i}$ e N_{h_i} sarebbe risultata sempre pari a ± 1 (anche se la covarianza, in media, è corretta) e avrebbe reso i risultati di più difficile interpretazione.

I risultati del confronto tra le stime del totale sono riportati nella tavola 1. La presentazione di detti risultati è stata effettuata a livello di area geografica, aggregando le stime relative ai 60 strati in 5 valori sintetici.

Dall'analisi della tavola 1 risulta evidente che lo stimatore $\hat{T}_1$ è caratterizzato da un MSE sensibilmente inferiore allo stimatore $\hat{T}$ per le variabili reddito e consumo (familiari); una apprezzabile riduzione di variabilità si verifica anche per la stima del totale della ricchezza, mentre non si rilevano differenze significative per quanto riguarda il risparmio. E' interessante notare inoltre che la riduzione di variabilità si presenta differenziata tra le varie aree geografiche; questo fenomeno è determinato non solo dal diverso grado di sovrarappresentazione delle UP di dimensioni maggiori all'interno delle

aree, ma anche dalla covarianza esistente tra N e $\hat{T}$ nelle UP, come è immediato verificare dall'esame della formula (5).

Infine, per quanto riguarda la modesta efficienza relativa di ${}_1\hat{T}$ rispetto a $\hat{T}$ per il risparmio, è opportuno ricordare l'elevata reticenza delle famiglie a fornire informazioni su questa variabile; tale fenomeno si presenta in misura abbastanza differenziata tra le UP e determina pertanto una notevole instabilità della stima della covarianza tra N_i e $\hat{T}_i$, con la conseguente perdita dell'efficienza relativa dello stimatore ${}_1\hat{T}$.

Confronto tra le stime del totale(*)

Reddito familiare	Var. rel. (1) x 10000	MSE. rel. (2) x 10000	D1 rel. (3) x 100
Area			
Nord-Ovest	28,98	53,28	25,09
Nord-Est	31,67	20,35	52,35
Centro	16,02	15,60	32,73
Sud	36,11	26,94	56,56
Isole	154,63	128,39	15,22
Totale	7,71	8,92	32,80
Consumo familiare	Var. rel. (1) x 10000	MSE. rel. (2) x 10000	D1 rel. (3) x 100
Area			
Nord-Ovest	31,63	31,44	59,23
Nord-Est	39,00	19,15	63,65
Centro	13,47	11,08	44,88
Sud	27,41	18,59	60,40
Isole	135,88	113,32	14,17
Totale	7,69	5,98	54,33
Ricchezza familiare	Var. rel. (1) x 10000	MSE. rel. (2) x 10000	D1 rel. (3) x 100
Area			
Nord-Ovest	106,85	220,35	10,87
Nord-Est	67,48	80,82	9,46
Centro	30,82	49,27	-4,89
Sud	143,47	167,35	32,87
Isole	309,74	248,02	10,86
Totale	22,32	31,29	15,42
Risparmio familiare	Var. rel. (1) x 10000	MSE. rel. (2) x 10000	D1 rel. (3) x 100
Area			
Nord-Ovest	200,82	430,59	9,24
Nord-Est	281,72	419,73	-11,17
Centro	236,24	548,36	-39,04
Sud	201,43	281,52	18,74
Isole	641,40	475,59	15,68
Totale	58,76	101,76	1,38

(*) Strati raggruppati in 12 ripartizioni geografiche x 5 classi di ampiezza demografica dei comuni

(1) Varianza relativa della stima del totale (1): $\hat{V}(\hat{T})/\hat{T}^2$

(2) Momento secondo dell'errore relativo del totale (3): $MSE(\hat{T})/\hat{T}^2$

(3) Stima di D_1 (formula (6)) in percentuale della varianza del totale: $D_1 \text{ rel.} = \hat{D}_1 / \hat{V}(\hat{T})$.

2. La stratificazione a posteriori

Nei campioni di modesta numerosità il collegamento con indagini molto più ampie può essere utile per verificare la rappresentatività del campione, per effettuare controlli sulla qualità delle informazioni rilevate e per migliorare l'efficienza delle stime attraverso procedimenti di post-stratificazione.

Al fine di valutare gli effetti che detti procedimenti producono, si supponga di disporre di due campioni (C_0 e C_1) estratti dalla stessa popolazione (ad esempio le famiglie italiane) ma caratterizzati da una diversa numerosità campionaria (n_0 e n_1 , $n_1 \gg n_0$); si supponga inoltre che i due campioni contengano rispettivamente due sottoinsiemi di informazioni, il primo costituito da variabili di struttura (quali età, sesso, condizione professionale), il secondo dagli altri dati rilevati (ad esempio reddito o consumo).

Sulla base delle ipotesi formulate è quindi possibile valutare l'opportunità di utilizzare le informazioni di struttura relative al campione di maggiore numerosità, calcolando in che modo la varianza di stime di totali si modifica ricorrendo alla post-stratificazione.

A questo scopo si confrontano due stimatori non distorti del totale di una variabile X , cioè lo stimatore $\hat{T}$, definito nella (1), e lo stimatore $_{(0/1)}\hat{T}$, che tiene conto delle informazioni rilevate con il campione più numeroso. Il deponente a sinistra di $\hat{T}$ indica che la stima, relativa a una variabile rilevata nel campione C_0 , viene effettuata sulla base delle informazioni di struttura del campione C_1 . Lo stimatore $_{(0/1)}\hat{T}$ si ottiene, dopo aver ripartito le osservazioni in sottoinsiemi per ogni UP e sulla base delle variabili di struttura, ponderando le medie di X per ogni sottoin-

sieme con i pesi che risultano dal campione di maggiore dimensione:

$$(7) \quad (O/1)\hat{T} = \sum_h^L M_h/m_h \sum_1^{m_h} N_{hi} (O/1)\bar{x}_{hi}$$

dove:

$$(O/1)\bar{x}_{hi} = 1/n_{hi} \sum_g^{G_{hi}} 1/n_{hig} \bar{x}_{hig}$$

$$\bar{x}_{hig} = 1/n_{hig} \sum_j^{n_{hig}} x_{higj}$$

Nel campionamento a due stadi stratificato al primo stadio con UP selezionate con probabilità costanti e senza reimmissione e con US estratte in ogni UP con uguale probabilità e con reimmissione 9/ la varianza della stima del totale definito nella (1) può essere scritta come

$$(8) \quad v(\hat{T}) = \sum_h^L M_h^2/m_h (M_h - m_h)/M_h S_{1h}^2 + \sum_h^L M_h/m_h \sum_1^{m_h} N_{hi}^2/n_{hi} v_{hi}(x)$$

dove:

$$S_{1h}^2 = 1/(M_h - 1) \sum_1^{M_h} (T_{hi} - T_h/M_h)^2$$

$$v_{hi}(x) = 1/N_{hi} \sum_j^{N_{hi}} (x_{hij} - T_{hi}/N_{hi})^2$$

mentre la varianza dello stimatore definito nella (7) è

$$(9) \quad v_{(0/1)\hat{T}} = \sum_h^L M_h^2/m_h(M_h-m_h)/M_h S_{ih}^2 + \sum_h^L M_h/m_h \sum_1^{M_h} N_{hi}^2 v_{(0/1)\bar{x}_{hi}}$$

dove per l'approssimazione di $v_{(0/1)\bar{x}_{hi}}$ si è fatto ricorso alla seguente espressione 10:

$$v_{(0/1)\bar{x}_{hi}} \approx 1/n_{hi} \cdot 1/N_{hi} \sum_g^{G_{hi}} N_{hig} v_{hig}(x) (1+(1-P_{hig})/1n_{hi}P_{hig}) \cdot$$

$$(1+(1-P_{hig})/n_{hi}P_{hig}) + 1/1n_{hi} \cdot 1/N_{hi} \sum_g^{G_{hi}} N_{hig} (T_{hig}/N_{hig} - T_{hi}/N_{hi})^2$$

$$v_{hig}(x) = 1/N_{hig} \sum_j^{N_{hig}} (x_{higj} - T_{hig}/N_{hig})^2$$

$$P_{hig} = N_{hig}/N_{hi}$$

La differenza (D_2) tra la varianza dei due stimatori (1) e (7), ovvero la differenza tra le formule (8) e (9), nell'ipotesi che le UP dei due campioni siano riferite alla stessa estrazione, è pari a 11/

$$(10) \quad D_2 = v(\hat{T}) - v_{(0/1)\hat{T}} = \sum_h^L M_h/m_h \sum_1^{M_h} N_{hi}^2 (1/n_{hi} v_{hi}(x) - v_{(0/1)\bar{x}_{hi}})$$

e può essere stimata mediante la seguente relazione, in analogia con quanto precisato prima della formula (6) :

$$(11) \quad \hat{D}_2 = \hat{v}(\hat{T}) - \hat{v}_{(0/1)\hat{T}} = \sum_h^L M_h^2/m_h^2 \sum_1^{m_h} N_{hi}^2 (1/n_{hi} \hat{v}_{hi}(x) - \hat{v}_{(0/1)\bar{x}_{hi}})$$

dove:

$$\hat{V}_{(0/1)\bar{x}_{hi}} = 1/n_{hi} \cdot 1/N_{hi} \sum_g^{G_{hi}} n_{hig} \hat{v}_{hig}(x) \cdot (1 + (1 - \hat{p}_{hig}) / (1/n_{hi} \hat{p}_{hig})) \cdot$$

$$\cdot (1 + (1 - \hat{p}_{hig}) / (n_{hi} \hat{p}_{hig})) + 1/n_{hi} s_{hi}^2$$

$$s_{hi}^2 = \hat{v}_{hi}(x) - 1/n_{hi} \sum_g^{G_{hi}} n_{hig} \hat{v}_{hig}(x)$$

$$\hat{p}_{hig} = n_{hig}/n_{hi}$$

dove $\hat{V}_{hig}$ e $\hat{v}_{hig}$ rappresentano rispettivamente stime corrette di V_{hig} e v_{hig} .

L'effetto post-stratificazione è stato stimato per le variabili reddito familiare e reddito individuale a livello di area geografica (Nord-Ovest, Nord-Est, Centro, Sud, Isole) e per il totale Italia. Per valutare quali parametri (variabili di struttura) risultano più significativi ai fini della riponderazione, l'analisi è stata effettuata ripartendo le osservazioni relative ad ogni UP in gruppi distinti per età, per titolo di studio, per condizione professionale, per numero di componenti e per numero di percettori di reddito. L'analisi, inoltre, è stata condotta per vari livelli teorici di numerosità del campione di riferimento (pari a 2,5 a 5 e a 10 volte il campione per l'indagine sui bilanci delle famiglie). Infine, per confrontare i risultati relativi al reddito familiare con i risultati relativi al reddito individuale, l'analisi è stata effettuata sui redditi dei soli capofamiglia; in tal caso infatti i campioni sono caratterizzati dalla stessa numerosità.

La varianza relativa della stima del totale della popolazione è risultata piuttosto insensibile al procedimento di post-stratificazione, per tutti i para-

metri e i livelli teorici di numerosità considerati. La maggiore riduzione di varianza si è verificata in seguito alla post-stratificazione per condizione professionale, sia per il reddito familiare che per il reddito individuale; tuttavia, per la modesta (cfr. tavola 2) entità di detta riduzione, non si ritiene di particolare importanza applicare correntemente procedimenti di riponderazione ai soli fini della riduzione di varianza delle stime del totale; detti procedimenti divengono invece notevolmente importanti per scopi particolari, come ad esempio per eliminare le distorsioni determinate dalla diversa probabilità di estrazione delle US o per confrontare due diversi campioni 12/.

Confronto tra le varianze delle stime del totale

! Reddito familiare(1) !	! Var. rel. (2) ! ! x 10000 !	! Var. rel. (3) ! ! x 10000 !	! D2 rel. (4) ! ! x 100 !
! Area !			
! Nord-Ovest !	26,02	25,10	3,53
! Nord-Est !	30,56	30,00	1,82
! Centro !	19,90	19,35	2,80
! Sud !	46,85	45,74	2,36
! Isole !	153,51	152,02	0,97
! Totale !	7,79	7,57	2,81
! Reddito individuale (1) !	! Var. rel. (2) ! ! x 10000 !	! Var. rel. (3) ! ! x 10000 !	! D2 rel. (4) ! ! x 100 !
! Area !			
! Nord-Ovest !	30,17	28,86	4,33
! Nord-Est !	37,33	36,05	3,41
! Centro !	12,43	11,37	8,61
! Sud !	60,11	57,72	3,97
! Isole !	116,80	114,34	2,10
! Totale !	8,58	8,21	4,23

- (1) Post-stratificazione per condizione professionale del capofamiglia ($n_1=10*n$)
- (2) Varianza relativa della stima del totale (1): $V(\hat{T})/\hat{T}^2$
- (3) Varianza relativa della stima del totale (7): $\hat{V}_{(0,1)}(\hat{T})/\hat{T}^2$
- (4) Riduzione di varianza in percentuale di $\hat{V}(\hat{T})$: $D2 \text{ rel.} = \hat{D}_2 / \hat{V}(\hat{T})$.

3. Conclusioni

Dall'analisi condotta sui dati relativi all'anno 1984, con i limiti che derivano dall'impossibilità di disporre, per gli anni precedenti, delle informazioni necessarie per la valutazione degli errori campionari e tenuto conto delle caratteristiche del piano di campionamento utilizzato, possono trarsi le seguenti conclusioni:

- lo stimatore $\hat{T}_1$ risulta maggiormente efficiente dello stimatore $\hat{T}$, nonostante i criteri di stratificazione tengano conto dell'ampiezza demografica dei comuni;
- l'uso di procedimenti di post-stratificazione, particolarmente importanti per eliminare distorsioni nelle stime di totali e/o valori medi, non sembra contribuire alla riduzione significativa della variabilità degli stimatori.

Infine, per quanto attiene alle stime pubblicate dalla Banca d'Italia (1985), si ritiene opportuno indicare che queste sono ottenute utilizzando uno stimatore del tipo (3), con opportuni pesi che tengono conto del numero di interviste assegnate e rientrate per ciascuno strato 13/; gli errori campionari relativi allo stimatore (3) possono pertanto essere considerati ragionevoli approssimazioni di quelli relativi alle stime pubblicate.

Al fine di agevolare l'analisi statistica dei risultati dell'indagine sui bilanci delle famiglie i dati della tavola 1 relativi a tale stimatore sono stati rielaborati e presentati (tavola 3) sotto forma di errori standard delle principali grandezze rilevate.

Tav. 3

**Valori medi e errori standard delle principali
variabili rilevate(*)**
(migliaia di lire)

! Variabile !	Stima del valore	Errore standard
! (1) !	! medio (2) !	! della stima !
! Reddito !	22.325	666,7
! Consumo !	15.339	375,1
! Risparmio !	2.194	221,3
! Ricchezza !	83.284	4658,7

(*) Indagine sui bilanci delle famiglie italiane
nell'anno 1984

(1) Stime riferite alle famiglie

(2) Stimatore $\hat{T}_1/N = \hat{T}/\hat{N}$

Dall'esame della tavola 3 risulta evidente che le stime dei valori medi ottenute con lo stimatore $\hat{T}_1$ si presentano molto prossime a quelle pubblicate (per il reddito familiare e la spesa per consumi le differenze sono rispettivamente pari a 0,35 per cento e 0,62 per cento); gli errori standard risultano invece superiori a quelli calcolati nell'ipotesi di campionamento casuale semplice.

In particolare, al livello di probabilità del 95 per cento, si ottengono i seguenti limiti (espressi in percentuale del valore medio) degli intervalli di confidenza: per il reddito familiare $\pm 5,8$ per cento contro $\pm 2,2$; per la spesa per consumi $\pm 4,8$ per cento contro $\pm 1,9$; risulta pertanto evidente che le informazioni relative a sottoinsiemi di modesta numerosità devono essere analizzate con maggiore cautela rispetto a quanto indicato in Banca d'Italia(1983).

Note

- 1/ Nel corso del lavoro si fa riferimento alla stima del totale e non alla stima della media ($\bar{X}$) le cui proprietà possono essere semplicemente ottenute in virtù della relazione $\bar{X} = \hat{T}/N$, dove N è la numerosità della popolazione.
- 2/ Hansen Hurwitz Madow (1953).
- 3/ Hansen Hurwitz Madow (1953) pag 336 vol. I, pag 194 vol. II.
- 4/ Ad esempio si vedano Istat(1978) e Istat(1985).
- 5/ In un campione casuale semplice la convergenza è dell'ordine $1/n$; Hansen et Al.(1953) pag 113 vol. II.
- 6/ Hansen et Al.(1953) pag 321 vol. I.
- 7/ Banca d'Italia (1983).
- 8/ Cochran (1977); Russo (1985).
- 9/ Il disegno campionario per l'indagine BI è in realtà quello descritto nel paragrafo 1. Tuttavia poiché la frazione di campionamento delle US è trascurabile, si può utilizzare la formula (8) come una ragionevole approssimazione della (2).
- 10/ Il valore di $V(\bar{X}_{(1/n)}^*)$ può essere calcolato esattamente ma l'espressione che ne risulta ha una forma particolarmente complessa. Si veda in proposito Stephan(1945).
- 11/ Tale differenza non necessariamente risulta positiva, anche nel caso che si disponga di informazioni di struttura relative all'universo. Hansen et Al. (1953) pag. 139 vol. II.
- 12/ Banca d'Italia(1985).
- 13/ Banca d'Italia (1983) pag 336.

APPENDICE

Tabella dei simboli utilizzati

h	indice di strato
i	indice di UP
g	indice di gruppo
j	indice di US
X_{higj}	valore del carattere X osservato sulla j-esima US nel g-esimo gruppo della i-esima UP dello h-esimo strato
X_{hij}	valore del carattere X osservato sulla j-esima US nella i-esima UP dell'h-esimo strato
T_{hig}	totale del carattere X nel gruppo g della i-esima UP dell'h-esimo strato
T_{hi}	totale del carattere X nella i-esima UP dello h-esimo strato
T_h	totale del carattere X nello strato h-esimo
T	totale del carattere X nella popolazione
n_{hig}	numero di osservazioni nel g-esimo gruppo dell'i-esima UP dell'h-esimo strato per il campione C_0
n_{hi}	numero di osservazioni nell'i-esima UP dell'h-esimo strato per il campione C_0
${}_1n_{hig}$	numero di osservazioni nel g-esimo gruppo dell'i-esima UP dell'h-esimo strato per il campione C_1
${}_1n_{hi}$	numero di osservazioni nell'i-esima UP dell'h-esimo strato per il campione C_1
N_{hig}	numero di US-universo nel g-esimo gruppo dell'i-esima UP dell'h-esimo strato
N_{hi}	numero di US-universo nell'i-esima UP dell'h-esimo strato

m_h	numero di UP-campione nell'h-esimo strato
M_h	numero di UP-universo nell'h-esimo strato
G_{hi}	numero di gruppi nella i-esima UP dello h-esimo strato
L	numero di strati

BIBLIOGRAFIA

- BANCA D'ITALIA (1983) - L'indagine campionaria sui bilanci delle famiglie italiane. Nota metodologica, "Bollettino Statistico" n. 3-4, Roma.
- BANCA D'ITALIA (1985) - I bilanci delle famiglie italiane nell'anno 1984, "Bollettino Statistico" n. 3-4, Roma.
- COCHRAN, W. G. (1977) - Sampling Techniques, New York, Wiley.
- HANSEN, M. H. - HURWITZ, W. N. - MADOW, W. G. (1953) - Sample Survey Methods and Theory, Vol. I e II, New York, Wiley.
- ISTAT (1978) - Rilevazione campionaria delle forze di lavoro, "Metodi e Norme", serie A, n. 15, Roma, Istat.
- ISTAT (1985) - Indagine sulle strutture ed i comportamenti familiari, Roma, Istat.
- RUSSO, A. (1985) - Un metodo di stima dell'effetto della stratificazione nei campioni complessi, "Quaderni di discussione", Roma, Istat.
- STEPHAN, F. F. (1945) - The Expected Value and the Variance of the Reciprocal and Other Negative Powers of a Positive Bernoullian Variate, "Annals Math. Stat.", XVI.

ELENCO DEI PIÙ RECENTI TEMI DI DISCUSSIONE (*)

- n. 73 — *Rischio e rendimento dei titoli a tasso fisso e a tasso variabile in un modello stocastico univariato*, di E. BARONE - R. CESARI (agosto 1986).
- n. 74 — *Gli strumenti per il sostegno pubblico dei carichi familiari: una valutazione quantitativa degli effetti redistributivi e degli oneri per la finanza pubblica*, di D. FRANCO - N. SARTOR (agosto 1986).
- n. 75 — *The Use of Italian Survey Data in the Analysis of the Formation of Inflation Expectations*, by I. VISCO (ottobre 1986).
- n. 76 — *Riflessioni e confronti in tema di separatezza tra banca e industria*, di R. PEPE (ottobre 1986).
- n. 77 — *L'internazionalizzazione del sistema bancario italiano. Una prospettiva di vigilanza*, di G. LANCIOTTI (ottobre 1986).
- n. 78 — *Determinazione del livello dei prezzi e politica «monetaria» in un'economia senza moneta*, di C. GIANNINI (novembre 1986).
- n. 79 — *Modifiche strutturali e tendenze dell'agricoltura italiana (1961-1982)*, di L. F. SIGNORINI - G. ZEN (novembre 1986).
- n. 80 — *Modello econometrico dell'economia italiana*. Vol. 1°: *struttura e proprietà*; Vol. 2°: *equazioni e note tecniche* (dicembre 1986).
- n. 81 — *Nuove valutazioni della capacità utilizzata in Italia*, di L. F. SIGNORINI (dicembre 1986).
- 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).

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

