



BANCA D'ITALIA
EUROSISTEMA

Supplementi al Bollettino Statistico

Note metodologiche

La misurazione dell'errore di stima nelle statistiche sui tassi di interesse bancari

Nuova serie

Anno XVII Numero 36 - 14 Giugno 2007

SUPPLEMENTI AL BOLLETTINO STATISTICO

Istituzioni finanziarie monetarie: banche e fondi comuni monetari (mensile)

Mercato finanziario (mensile)

Finanza pubblica (mensile)

Bilancia dei pagamenti (mensile)

Aggregati monetari e creditizi dell'area dell'euro: le componenti italiane (mensile)

Conti finanziari (trimestrale)

Sistema dei pagamenti (semestrale)

Statistiche di finanza pubblica nei paesi dell'Unione europea (annuale)

Debito delle Amministrazioni locali (annuale)

Indagini campionarie (periodicità variabile)

Note metodologiche (periodicità variabile)

Tutti i supplementi sono disponibili sul sito Internet della Banca d'Italia: www.bancaditalia.it

Eventuali chiarimenti sui dati contenuti in questa pubblicazione possono essere richiesti via e-mail all'indirizzo statistiche@bancaditalia.it

INDICE

	pag.
1. Introduzione e principali conclusioni	5
2. L'errore di stima nelle statistiche sui tassi di interesse bancari	5
2.1 Lo stimatore dei tassi di interesse bancari come stimatore di un quoziente	7
2.2 L'applicazione del metodo Jackknife alla misurazione dell'errore di stima nelle statistiche sui tassi di interesse bancari	8
3. Risultati	9
Bibliografia	12
Appendice	13

LA MISURAZIONE DELL'ERRORE DI STIMA NEI TASSI DI INTERESSE BANCARI

Massimiliano Stacchini¹

1. Introduzione e principali conclusioni

Il Regolamento BCE/2001/18, che disciplina la raccolta delle informazioni statistiche relative ai tassi bancari armonizzati per l'area dell'euro prevede la possibilità di raccogliere informazioni su base campionaria ma stabilisce che, dai dati riferiti al gennaio 2007, il campione debba soddisfare un requisito di rappresentatività in base al quale l'errore massimo per i tassi di interesse sulle nuove operazioni, relativo a una media su tutte le categorie di strumenti, non deve superare 10 punti base, ad un livello di confidenza del 90 per cento (par. 10, Annex I).

Nel caso questa condizione non fosse soddisfatta, si dovrebbe procedere ad un allargamento del campione. In vista di questa scadenza, si è verificato se il campione stia attualmente soddisfacendo il requisito di rappresentatività.

Questa nota riassume la metodologia impiegata per misurare l'errore di stima della statistica $\hat{\theta}_{mir}$ utilizzata per stimare i tassi armonizzati. L'errore è stimato sulla base dei dati relativi al periodo gennaio 2005 – ottobre 2006 (ultimo dato disponibile al momento dell'esercizio).

I risultati ottenuti confermano che il campione attualmente utilizzato (121 banche) soddisfa il requisito in via di introduzione. Questa conclusione vale per ciascuna delle tre metodologie utilizzate per ricavare la varianza dello stimatore.

2. L'errore di stima nelle statistiche sui tassi di interesse bancari

Per ciascuno dei fenomeni previsti dal Regolamento 2001/18 la Banca d'Italia raccoglie le informazioni sui tassi di interesse bancari definite dalla statistica

$$\hat{\theta}_{mir} = \frac{\sum_i^n r_i w_i}{\sum_i^n w_i} = \frac{\sum_h^L \sum_i^{n_h} r_{hi} w_{hi}}{\sum_h^L \sum_i^{n_h} w_{hi}} \quad \text{dove}$$

$\hat{\theta}_{mir}$ è uno stimatore del parametro θ , il tasso di interesse per l'intera popolazione oggetto dell'indagine (quasi 800 banche); r_i è il tasso d'interesse della banca i -esima; w_i è l'importo segnalato dalla banca i -esima; r_{hi} è il tasso d'interesse della banca i -esima nello strato h ; w_{hi} è l'importo della banca i -esima

¹ Banca d'Italia, Servizio Studi. Desidero ringraziare Rosaria Buzzi per l'aiuto nella raccolta delle statistiche sui tassi di interesse bancari e Stefano Iezzi, Paola Battipaglia e Matteo Piazza per gli utili commenti al lavoro.

nello strato h ; infine n e n_h indicano, rispettivamente, la numerosità del campione e nello strato h ($h = 1, \dots, L$).

L'errore di stima (D) è definito come

$$D = Z_{\alpha/2} \sqrt{V(\hat{\theta}_{mir})} \approx Z_{\alpha/2} \sqrt{v(\hat{\theta}_{mir})} \quad (1)$$

dove $Z_{\alpha/2}$ è il $(1-\alpha/2)*100$ -esimo percentile della distribuzione appropriata, $V(\hat{\theta}_{mir})$ è la varianza di $\hat{\theta}_{mir}$, e $v(\hat{\theta}_{mir})$ un suo stimatore.

Nella costruzione di $\hat{\theta}_{mir}$, i tassi bancari segnalati dalla banca i -esima del campione sono ponderati attraverso l'importo segnalato dalla banca. In generale, la ponderazione delle informazioni definisce in modo appropriato la "rappresentatività" di una unità campionaria, conferendo allo stimatore la proprietà della *correttezza*. In assenza della ponderazione, le unità campionarie, alle quali il disegno di campionamento attribuisce 'gradi' differenti di rappresentatività, entrerebbero nella statistica con peso uguale, distorcendo lo stimatore.

Allo stesso tempo, la ponderazione delle informazioni accresce la varianza dello stimatore e dell'errore di stima, per la variabilità dei coefficienti di ponderazione tra le banche.

Il problema della misurazione dell'errore di stima, in presenza di coefficienti di ponderazione, è risolto dalla letteratura in due modi.

Un approccio alla base della prima metodologia utilizzata in questa nota riconosce nell'importo associato al tasso di interesse una "caratteristica" dell'unità campionaria, al pari del tasso di interesse. L'*errore di stima* viene così a dipendere dalla varianza del rapporto tra due "caratteristiche" dell'unità campionaria: il flusso di interessi pagati o ricevuti dalla banca, da un lato, il suo volume di operatività, dall'altro.

Una diversa strada è seguita dai metodi per la *replicazione del campione* che forniscono espressioni, non derivate analiticamente, per la varianza degli stimatori. Tra questi metodi rientra la procedura *Jackknife* (Quenouille, 1956), impiegata per la seconda misurazione dell'errore di stima.

La terza metodologia utilizzata è quella proposta da Battipaglia e Bolognesi (2003) nella fase di avvio della rilevazione dei tassi d'interesse armonizzati e costituisce quindi un naturale *benchmark* di controllo dei risultati ottenuti.

Di seguito sono presentati i due approcci applicati *ex novo* in questo lavoro (per riferimenti sull'approccio di Battipaglia e Bolognesi si rimanda al loro lavoro del 2003) e i risultati relativi ai tre metodi.

2.1 Lo stimatore dei tassi di interesse bancari come stimatore di un quoziente

Si supponga di avere una popolazione composta da h strati, dove Y_{hi} e X_{hi} sono le modalità assunte da due “caratteristiche” Y e X dell’unità i -esima nello strato h . Si definisca il quoziente

$R = \frac{Y}{X}$, dove $Y = \sum_{h=1}^L \sum_{i=1}^{N_h} Y_{hi}$ e $X = \sum_{h=1}^L \sum_{i=1}^{N_h} X_{hi}$ sono i totali delle caratteristiche Y e X , riferiti all’intera

popolazione. Si definisca inoltre lo stimatore $\hat{R} = \frac{\hat{Y}}{\hat{X}}$, dove \hat{Y} e \hat{X} sono *stimatori corretti*, mediante

l’informazione campionaria, dei totali delle caratteristiche Y e X . Ponendo $\hat{Y} = \sum_{h=1}^L \sum_{i=1}^{n_h} y_{hi}$,

$\hat{X} = \sum_{h=1}^L \sum_{i=1}^{n_h} x_{hi}$, $y_{hi} \equiv r_{hi} * w_{hi}$ e $x_{hi} \equiv w_{hi}$ si ottiene $\hat{R} = \frac{\hat{Y}}{\hat{X}} \equiv \hat{\theta}_{mir}$.

In questo schema $\hat{\theta}_{mir}$ è uno stimatore del quoziente tra gli stimatori dei totali Y e X .² Da tale identità discende l’espressione approssimata della varianza:

$$v(\hat{R}) = v(\hat{\theta}_{mir}).$$

Lo stimatore \hat{R} del quoziente tra due totali è una funzione *non lineare* nei due totali Y e X (Rao, 2000). La non linearità dello stimatore \hat{R} rende difficoltosa la derivazione analitica di un’espressione per la varianza. Un risultato utile è ottenibile con la tecnica della *linearizzazione della varianza*³.

Se il disegno campionario prevede la stratificazione, la trasformata di Woodruff fornisce l’espressione:

$$v(\hat{R}) = \frac{\sum_i n_h (1 - f_h)}{\hat{X}^2} \frac{\sum_1^n [(y_{hi} - \bar{y}_h) - \hat{R}(x_{hi} - \bar{x}_h)]^2}{n_h - 1}$$

dove \bar{y}_h è la media dei valori di y_{hi} nello strato h , \bar{x}_h è la media dei valori x_{hi} nello strato h ed f_h è la frazione sondata di strato. La formula riportata per il campionamento stratificato è quella del *combined ratio estimator* (Cicchitelli, 1993). Tale stimatore è *non distorto* e su di esso si orienta la scelta quando la numerosità campionaria dei singoli strati non è elevata⁴ (Rao, 2000).

² Questa interpretazione è suggerita dal Manual on MFI interest rate statistics (BCE 2003): “It is important to note that in the case of a weighted average the variance of a ratio would have to be calculated”.

³ In base a questo approccio, stimatori non lineari sono espressi come funzione di stimatori lineari. Quindi si utilizza il ‘delta method’ per ottenere l’espressione per la varianza della varianza dello stimatore.

⁴ Per la stima di quozienti R nell’ambito di disegni campionari stratificati la letteratura offre due strumenti: il *separated estimator* e il *combined estimator*. Il primo assume che il campione possa stimare accuratamente i parziali di strato R_h . Il secondo stimatore, non stima i parziali di strato, pur tenendo conto del carattere stratificato del disegno campionario. Il *separated estimator* è più efficiente ma è distorto se le stime parziali R_h sono errate. Se le numerosità campionarie non sono

2.2 L'applicazione del metodo Jackknife alla misurazione dell'errore di stima nelle statistiche sui tassi di interesse bancari

Un secondo modo di stimare la varianza di $\hat{\theta}_{mir}$ è fornito dai metodi per la *replicazione del campione*. Queste tecniche consentono di estrarre informazione dal campione, senza il filtro di funzioni analiticamente derivate. L'idea che accomuna i metodi per la *replicazione* del campione, tra cui il metodo *Jackknife* (JK) utilizzato in questo lavoro⁵, è di derivare una espressione per la varianza dello stimatore $\hat{\theta}_{mir}$, considerando un secondo stimatore $\hat{\theta}_{jack}$ del parametro θ , con le seguenti caratteristiche:

$$v(\hat{\theta}_{jack}) \cong v(\hat{\theta}_{mir}), \quad v(\hat{\theta}_{jack}) \text{ nota.}$$

Il problema della stima della varianza di $\hat{\theta}_{mir}$ è risolto indirettamente attraverso la conoscenza di $v(\hat{\theta}_{jack})$. Un'espressione per $v(\hat{\theta}_{jack})$ è ricavata dalle informazioni che si ottengono dalle *replicazioni* del campione originario⁶. La possibilità di generare delle *replicazioni*, utilizzabili come base informativa per stimare la varianza, è condizionata alla validità del *Glivenko-Kantelli theorem*:

$$Pr\left\{\lim_{n \rightarrow \infty} \text{diff}_n = 0\right\} = 1 \quad \text{dove} \quad \text{diff}_n = \sup_{-\infty < x < +\infty} |F_{n(x)} - F(x)|$$

In altri termini si assume che il campione $F_{n(x)}$ possa approssimare $F(x)$ e che pertanto possa essere utilizzato quale diretta espressione della popolazione oggetto di studio. Nel caso della rilevazione armonizzata, l'ipotesi di validità del *Glivenko-Kantelli theorem* è giustificata dall'elevata copertura del campione (superiore all'80 per cento degli stock rilevanti per la segnalazione armonizzata). Wolter (1985) e Cicchitelli (1993) discutono l'applicazione della tecnica JK a *disegni campionari stratificati, con unità estratte con probabilità variabili*. Altre applicazioni, per piani di campionamento stratificati a *più stadi*, sono contenuti in Yung e Rao (1996).

Nel caso di un disegno stratificato l'espressione utile è:

$$v(\theta_{jack}) = \sum_{h=1}^L (1-f_h) \frac{n_h - 1}{n_h} \sum_{l=1}^{n_h} (\hat{\theta}_{hl} - \hat{\theta}_h)^2$$

elevate nei singoli strati, la scelta cade sul *combined estimator* (Cicchitelli, 1993; Rao 2000). Il campione utilizzato per la rilevazione armonizzata ha, per gli strati non autorappresentativi, numerosità pari a $n1=34, n2=12, n3=13$.

⁵ Un'altra tecnica che utilizza la replicazione dei campioni è il *bootstrap*. Tuttavia la letteratura non ha al momento sviluppato adattamenti di questa tecnica a disegni campionari stratificati (Wolter 1985, Sarndal et al. 1992).

⁶ Le replicazioni sono costituite da sottocampioni ottenuti omettendo un certo numero di osservazioni dal campione originario. Un risultato della letteratura, applicato in questo lavoro, è quello per cui la precisione dello stimatore è massima quando le replicazioni del campione sono costituite da sottocampioni di $n-1$ unità.

dove $\hat{\theta}_{hi}$ è lo stimatore di θ che ha la stessa forma funzionale di $\hat{\theta}_{mir}$, ottenuto omettendo dal campione

l'osservazione i -esima nello strato h e in cui $\hat{\theta}_h = \sum_{i=1}^{n_h} \frac{\hat{\theta}_{hi}}{n_h}$.

3. Risultati

In questo paragrafo, si riportano i risultati ottenuti applicando le due metodologie descritte e la metodologia di misurazione dell'errore proposta in Battipaglia e Bolognesi (2003). In quest'ultimo caso, nelle stime preliminari all'avvio della rilevazione sui tassi armonizzati la varianza dello stimatore (dello strato non autorappresentativo) era stata approssimata combinando per singola forma tecnica, secondo la formula della varianza del prodotto di due variabili aleatorie, la stima della varianza campionaria del tasso di interesse medio con la stima della variabilità intertemporale della frazione sondata, espressa in termini di importi per le nuove operazioni. I risultati sono riportati nella tavola 1 e nelle figure 1-4 (gli acronimi *ratio*, *jack*, *bb* indicano i risultati ottenuti rispettivamente con la linearizzazione della varianza del *ratio estimator*, con la tecnica *Jackknife* e con il metodo seguito da Battipaglia et al).

L'errore di stima della statistica $\hat{\theta}_{mir}$, al livello di probabilità del 10 per cento è fornito dalla (1) dove si è posto $Z_{\alpha/2} = 1.645^7$. Sono state utilizzate 52.892 osservazioni *banca-mese*, comprendenti dati su tassi bancari e importi, segnalati nel periodo gennaio 2005 – ottobre 2006 dalle 121 banche del campione su ciascun fenomeno. La natura dell'errore è *point-in-time*, essendo definita per ciascuno dei 22 periodi dell'arco temporale considerato e per 28 dei 29 tassi di interesse sulle nuove operazioni previsti dal Regolamento ECB2001/18 (per un tasso, rilevato in modo censuario dal momento che un'unica banca - inclusa nel campione - è operativa in quel segmento di mercato, l'errore di stima non è definito).

Nella (1), In luogo di $v(\hat{\theta}_{mir})$, sono state utilizzate le posizioni $v(\hat{R})$ e $v(\hat{\theta}_{jack})$, discusse nei paragrafi precedenti⁸. Si sono misurati 1232 errori *point-in-time D*, (616 dal *ratio estimator*, 616 con la tecnica *Jackknife*). La varianza stimata da Battipaglia e Bolognesi è di natura "intertemporale": il loro metodo fornisce 28 errori, rappresentativi, per ciascun tasso, dell'intero periodo di osservazione

⁷ Nella formula (1) è impiegato il centile della *t di Student*. Secondo Cicchitelli (1993) "Per piani di campionamento diversi da quello casuale semplice [...] si deve fare affidamento in gran parte sui risultati delle simulazioni condotte da Frankel (1971), Bean (1975), Campbell e Meyer (1978), e riguardanti anche stimatori rappresentati da funzioni campionarie complesse. Gli schemi che più frequentemente sono esaminati prevedono la stratificazione delle unità primarie [...]. Ne discende che gli intervalli di confidenza vengono costruiti in modo naturale sulla base dei centili della distribuzione *t di Student*."

⁸ Operativamente la frazione sondata f_h delle $v(\hat{R})$ e $v(\hat{\theta}_{jack})$, è stata parametrizzata, per ciascuno strato, con il rapporto tra il numero di banche del campione e quello delle banche della popolazione (di strato), ovvero $f_h = 34/428 = 0,08$ se $h=1$, $f_h = 12/151 = 0,08$ se $h = 2$, $f_h = 13/146 = 0,08$ se $h = 3$, $f_h = 64/64 = 1$ se $h = 4$. I parametri f_h sono fissi, ossia non variano quando una banca del campione segnala importi nulli in una certa categoria e in un dato periodo. In tal modo, le banche del campione che segnalano importi nulli risultano rappresentative di quella frazione di popolazione di banche, non

Aggregando i risultati relativi ai 22 periodi e a tutti i fenomeni osservati, emerge che **l'errore sui tassi bancari è ampiamente al di sotto dei 10 punti base, soglia massima stabilita dal Regolamento**. Le tre tecniche impiegate forniscono risultati simili: l'errore stimato (errore medio) è pari, rispettivamente, a 0,049; 0,042; 0,042. La stima è inferiore se gli errori sui singoli fenomeni sono ponderati in base alla rilevanza dello stesso (errore medio ponderato). In questo caso l'errore è pari, rispettivamente, a 0,024; 0,024; 0,028. I valori inferiori si spiegano con la maggior rilevanza data ai modesti errori di stima dei fenomeni che hanno con volumi più grandi sulle nuove operazioni. L'errore di stima è più elevato per i tassi sui prestiti (0,064; 0,055; 0,053), che per i tassi sui depositi (0,020; 0,018; 0,022). La differenza dell'errore potrebbe essere guidata dalla differenza nei "livelli" dei tassi di interesse tra i due gruppi di fenomeni.

Tavola 1. Accuratezza delle statistiche sui tassi d'interesse

(errore di stima in punti percentuali; periodo: gennaio 2005 - ottobre 2006)

Tassi sulle nuove operazioni previste dal Regolamento 2001/18			
	ratio	jack	bb
errore medio	0.049	0.042	0.042
errore medio ponderato	0.024	0.024	0.028
errore medio sui prestiti	0.064	0.055	0.053
errore medio sui depositi	0.020	0.018	0.022
Tassi sulle nuove operazioni <i>stricto sensu</i>*			
	ratio	jack	bb
errore medio	0.052	0.043	0.043
errore medio ponderato	0.027	0.023	0.024

**Sono esclusi i prestiti in conto corrente, i depositi a vista e rimborsabili con preavviso, per i quali sono rilevate le sole consistenze del periodo.*

Altre aggregazioni forniscono l'errore di stima per i tassi sulle nuove operazioni *in senso stretto* (ottenute escludendo dai *new business* classificati dal Regolamento 2001/18 i prestiti in conto corrente, i depositi a vista e rimborsabili con preavviso). I risultati sono simili ai precedenti.

Dall'analisi dei risultati disaggregati a livello di fenomeno (figure 1-2) emerge che ciascun tasso rilevato ha un errore inferiore ai 10 punti base, ad eccezione del tasso sui prestiti per altri scopi e sul credito al consumo (entro 1 anno). Nel primo caso, due tecniche evidenziano un errore superiore a 30 punti base; nel secondo caso, l'errore è pari a circa 12 punti base. Il peso di queste categorie non supera l'1 per cento del totale delle nuove operazioni della rilevazione armonizzata; le stesse categorie hanno circa il 3 per cento del totale delle nuove operazioni *in senso stretto*.

sondata, segnalante importi nulli in una certa categoria e in un dato periodo.

Nelle figure 3-4 l'intervallo di confidenza misurato per l'Italia è confrontato con la banda ammessa dal Regolamento 2001/18. Nelle stesse figure si riporta il tasso di interesse sul fenomeno. La forma *point-in-time*, riferita alla singola data dell'intervallo, permette di seguire l'andamento dell'errore di stima nel tempo. Il confronto è presentato per il tasso sui prestiti in conto corrente alle imprese e quello sui depositi a scadenza delle famiglie, fenomeni rilevanti nel sistema bancario italiano, ma è stato effettuato per ciascuno dei tassi bancari analizzati nel lavoro. Osservando le figure si vede come l'errore di stima sia stabile e inferiore ai 10 punti base nell'intero arco temporale. Ciò vale anche per la fase interessata dal rialzo del tasso di riferimento per le operazioni di rifinanziamento. La stabilità, nel tempo, degli errori di stima riportati dalle figure 1-2, è stata verificata anche per i tassi di interesse il cui errore *point-in-time* non è illustrato nei grafici.

Bibliografia

- Battipaglia P. and F. Bolognesi, (2003), The Harmonization of European Statistics on Bank Interest Rates and the Methodology Adopted by Italy, *Supplements to the Statistical Bulletin*, Bank of Italy.
- European Central Bank, (2003), *Manual on MFI Interest Rate Statistics*.
- Cicchitelli G., Herzel A. and G. Montanari, (1993), *Il campionamento statistico*, Il Mulino.
- Cochran W.G., (1977), *Sampling Techniques*, John Wiley & Sons.
- Quenouille M.H., (1956), Notes on bias in estimation, *Biometrika*.
- Rao P., (2000), *Sampling Methodologies*, Chapman & Hall.
- Sarndal C.E., Swensson B., Wretman J., (1992), *Model Assisted Survey Sampling*, Springer-Verlag.
- Wolter K., (1985), *Introduction to Variance Estimation*, Springer-Verlag.

Appendice

Figura 1. Accuratezza delle statistiche sui tassi di interesse attivi

(errore di stima in punti percentuali; periodo: gennaio 2005 - ottobre 2006)

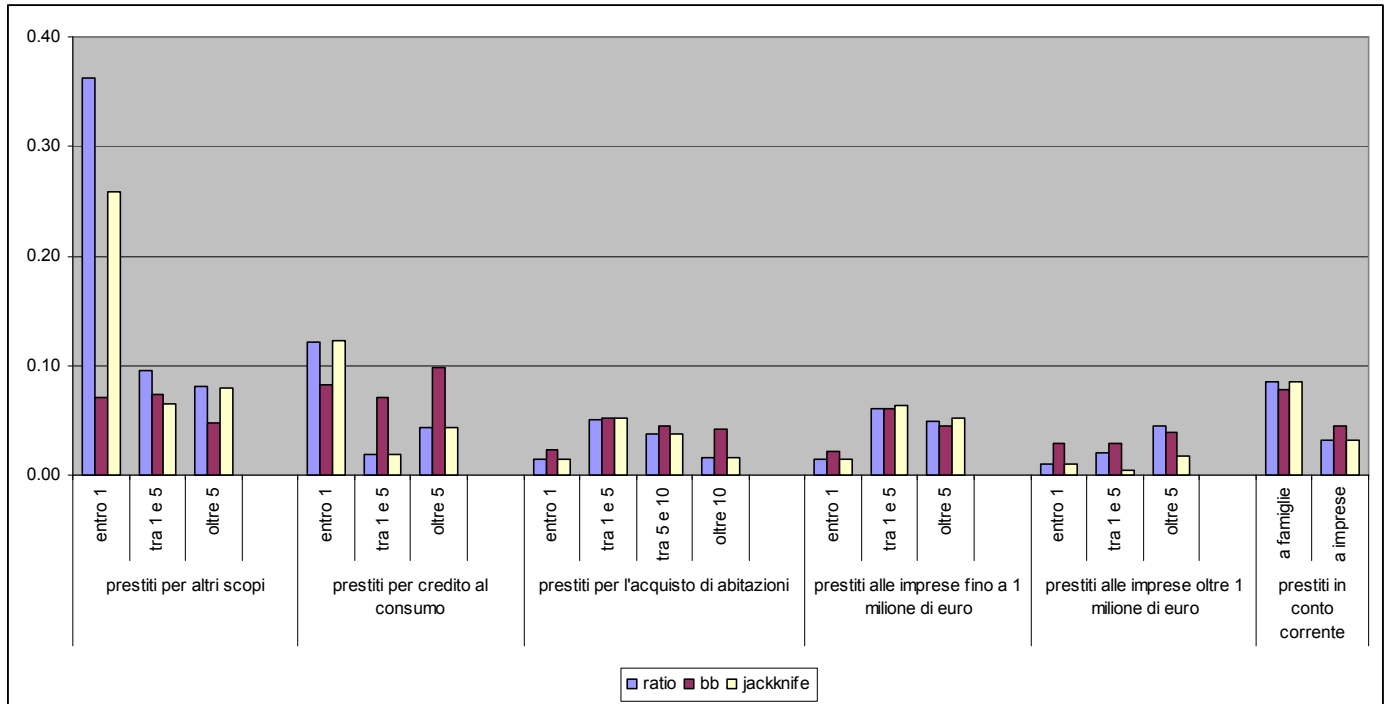


Figura 2. Accuratezza delle statistiche sui tassi di interesse passivi

(errore di stima in punti percentuali; periodo: gennaio 2005 - ottobre 2006)

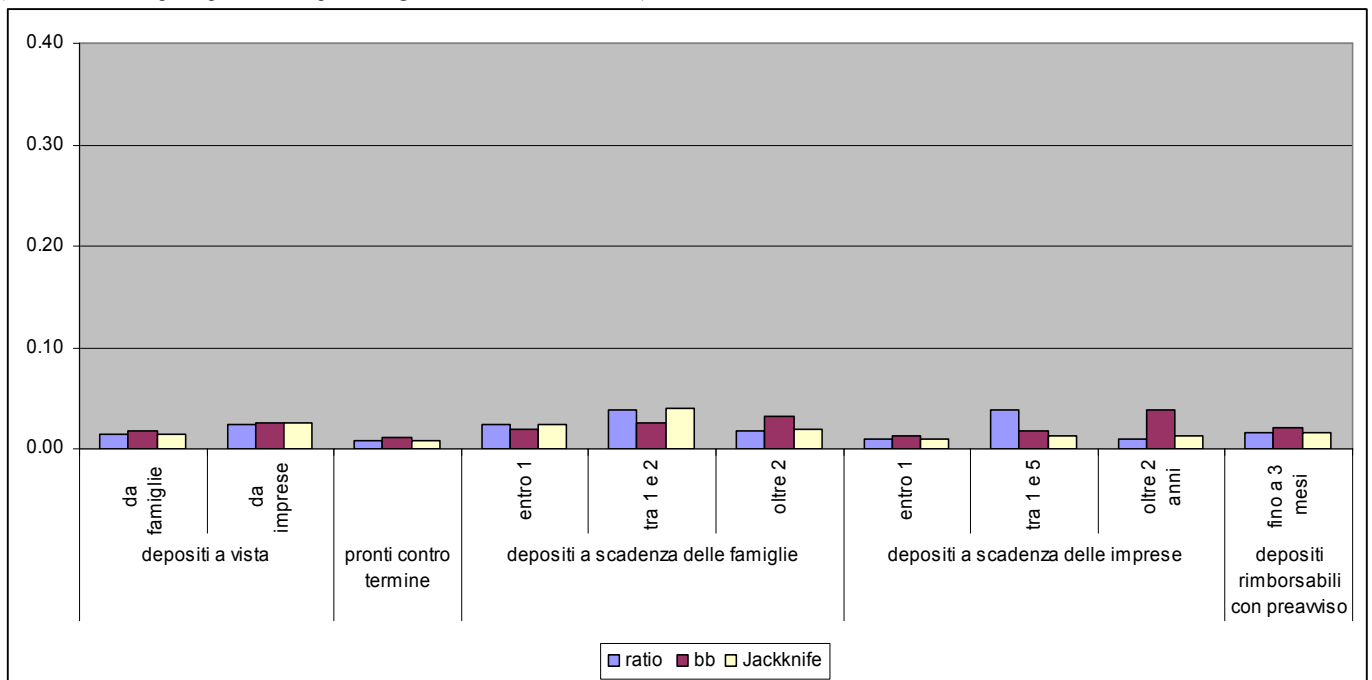


Figura 3. Accuratezza delle statistiche *point-in-time* sui tassi bancari: intervallo di confidenza ammesso dal Regolamento e misurato per l'Italia.

(le barre delimitano l'intervallo di confidenza ammesso dal Regolamento ECB 2001/18; la superficie in grigio quello per l'Italia sul fenomeno indicato; periodo: gennaio 2005 - ottobre 2006)

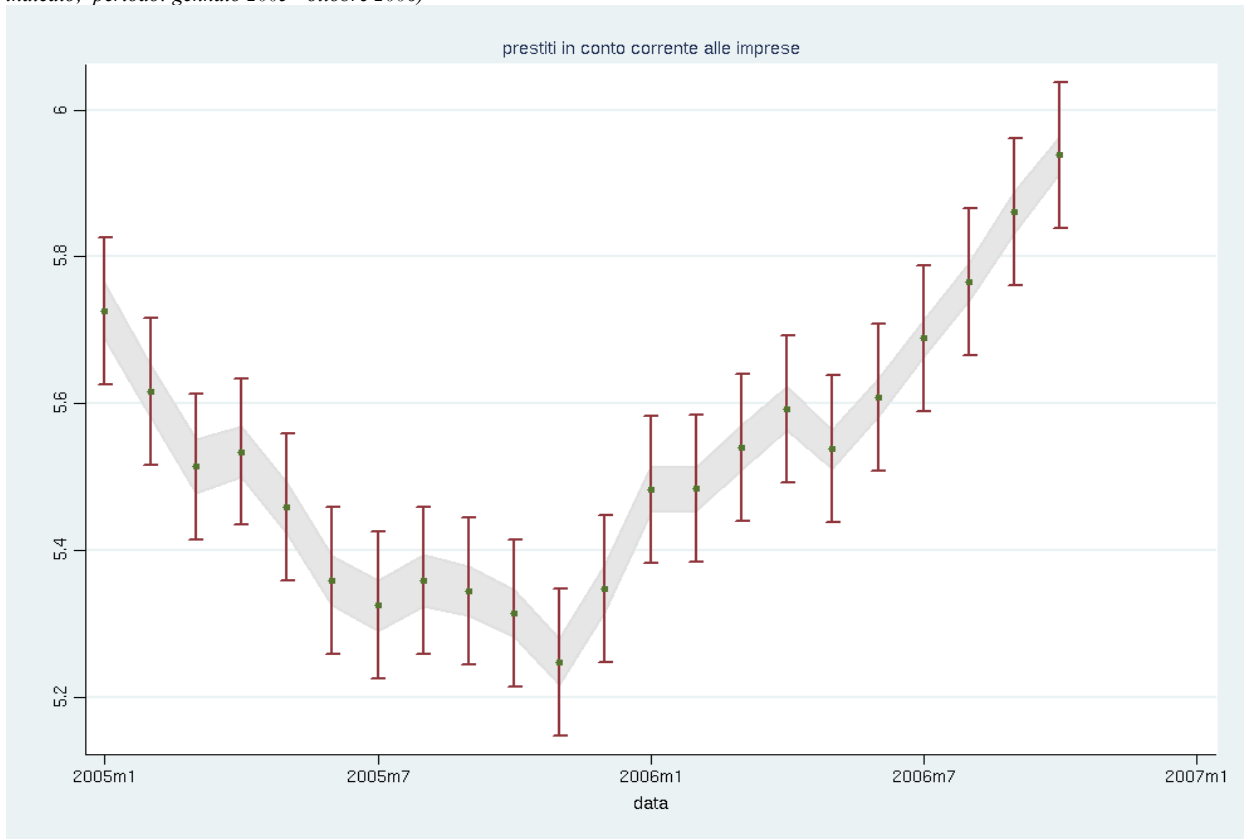
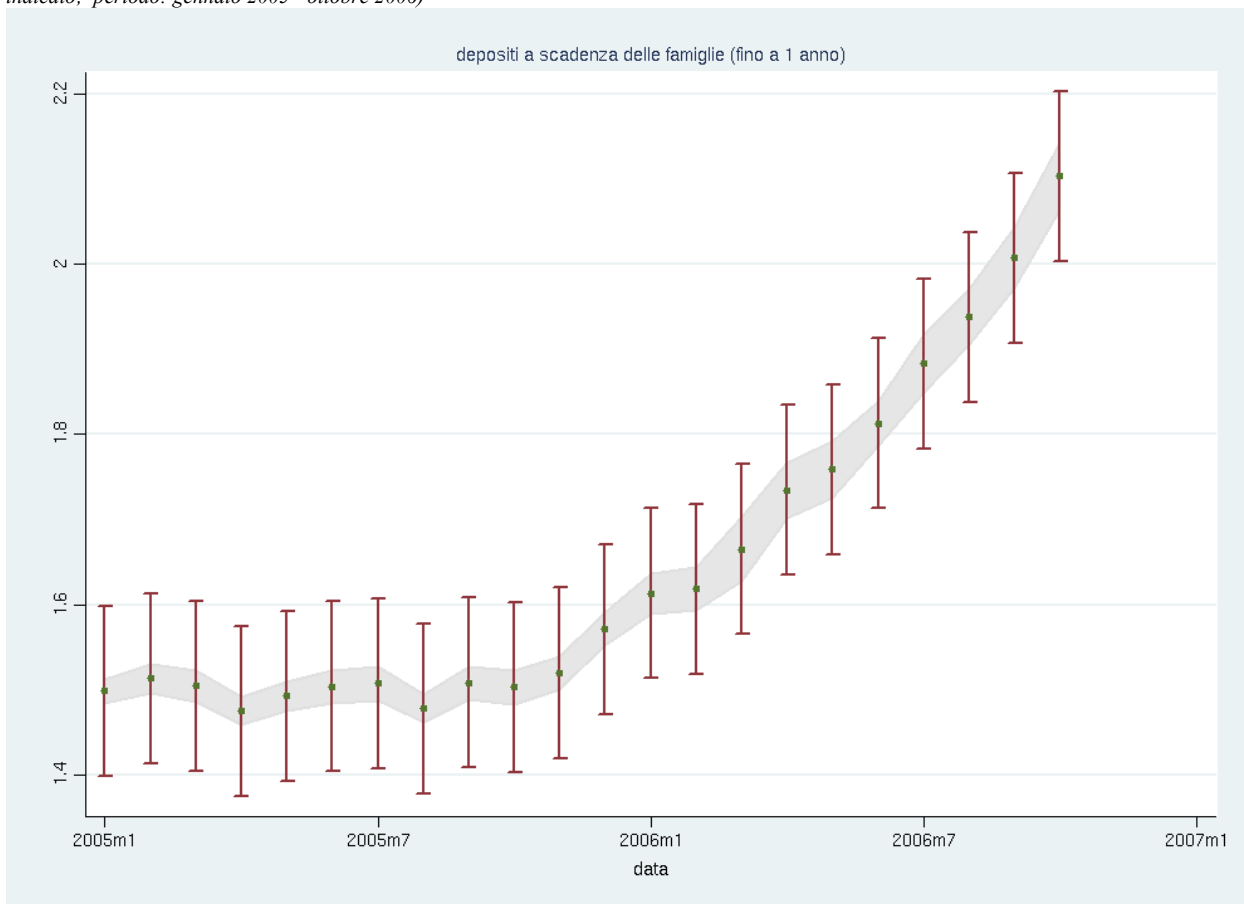


Figura 4. Accuratezza delle statistiche *point-in-time* sui tassi bancari: intervallo di confidenza ammesso dal Regolamento e misurato per l'Italia.

(le barre delimitano l'intervallo di confidenza ammesso dal Regolamento ECB 2001/18; la superficie in grigio quello per l'Italia sul fenomeno indicato; periodo: gennaio 2005 - ottobre 2006)



Stampa su carta riciclata