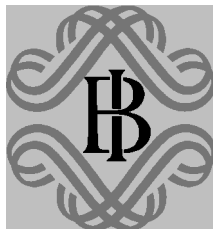


**BANCA D'ITALIA**

**Supplementi al Bollettino Statistico**  
**Note metodologiche e informazioni statistiche**

**L'utilizzo del *Selective editing* per il controllo  
di qualità delle statistiche bancarie**



**Nuova serie**

**Anno XIV Numero 29 - 24 Maggio 2004**

## **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)

**Note metodologiche e informazioni statistiche** (periodicità variabile)

*Tutti i supplementi sono disponibili sul sito Internet della Banca d'Italia : [www.bancaditalia.it](http://www.bancaditalia.it)*

*Eventuali chiarimenti sui dati contenuti in questa pubblicazione possono essere richiesti via e-mail all'indirizzo [statistiche@bancaditalia.it](mailto:statistiche@bancaditalia.it)*

# L'UTILIZZO DEL *SELECTIVE EDITING* PER IL CONTROLLO DI QUALITÀ DELLE STATISTICHE BANCARIE

P. Battipaglia, G. Bruno, F. Farabullini, C. Mahlknecht<sup>1</sup>

## 1. Introduzione

Una delle attività più impegnative nella produzione di statistiche per l'analisi economica è rappresentata dal controllo di qualità dei dati, soprattutto quando le serie sono numerose e a elevata frequenza. In queste condizioni risulta particolarmente difficile combinare i requisiti di tempestività e di qualità richiesti dagli utenti.

Il tema è qui affrontato con un'applicazione alle statistiche bancarie che la Banca d'Italia deve produrre mensilmente per la Banca centrale europea (BCE). Il *trade-off* tra qualità e tempestività assume particolare rilevanza, in quanto per il controllo e la produzione delle informazioni aggregate è disponibile un solo giorno lavorativo. In questo contesto è necessaria una procedura per la verifica di qualità dei dati che sia efficiente, affidabile, flessibile e con elevate prestazioni statistiche e informatiche.

In questo lavoro si sperimentano tecniche che combinano l'utilizzo di modelli autoregressivi a media mobile (ARIMA) con strategie di *selective editing*; queste ultime consentono di stabilire una graduatoria degli *outliers* individuati dai modelli ARIMA, in modo da rendere più efficiente l'attività di controllo. La graduatoria è costruita in base a una stima dell'impatto che i risultati della verifica degli *outliers* potrebbero avere su alcuni aggregati o indicatori significativi.

L'esercizio viene svolto utilizzando le 17 serie che compongono l'aggregato "contropartite della moneta M3"; tali serie riguardano gli impieghi concessi dalle banche italiane al settore detentore di moneta, nonché i titoli in portafoglio delle stesse banche.

Il lavoro è diviso in cinque parti. Nel paragrafo che segue si passano in rassegna le principali metodologie statistiche per il controllo di qualità dei dati; nella terza parte viene introdotta l'applicazione alle statistiche bancarie; la quarta sezione illustra la strategia proposta; l'ultima parte riporta i principali risultati e delinea i possibili sviluppi di ricerca.

---

<sup>1</sup> Banca d'Italia, Servizio Studi. Le opinioni espresse in questo lavoro non impegnano la responsabilità della Banca d'Italia.

## 2. Tecniche per l'identificazione e il trattamento dei valori anomali

L'identificazione di valori anomali rappresenta una fase cruciale nel controllo di qualità dei dati, che si completa con la conferma dei valori osservati o con la sostituzione di questi con dati rettificati o stimati. Le metodologie proposte, in letteratura e nelle applicazioni, sono numerose. Per una categorizzazione essenziale, si possono distinguere due approcci alternativi: quello parametrico individua un modello generatore delle osservazioni e ne sfrutta i valori caratteristici, quello non parametrico utilizza l'analisi empirica dei dati, prescindendo da aspettative basate su modelli teorici.

L'approccio parametrico, utilizzando le probabilità di appartenenza dei valori osservati alla distribuzione ipotizzata, consente di impostare *routines* automatiche che individuano i valori estremi in funzione di prefissati livelli di significatività, ed eventualmente li sostituiscono con valori stimati (cfr. ad esempio Alwan e Roberts, 1988). Le tecniche non parametriche – che spaziano dalle rappresentazioni grafiche, all'analisi delle distanze, ai sistemi di intelligenza artificiale – hanno il vantaggio di una maggiore flessibilità e di una maggiore capacità di adattamento a relazioni non lineari (cfr. Landenna e Marasini, 1990, Cheng e Titterington, 1994); esse, tuttavia, non consentono una rappresentazione sistematica utilizzabile per l'interpretazione e la previsione dell'andamento dei fenomeni e delle relazioni tra variabili.

Nelle procedure di validazione dei dati, all'individuazione dei valori anomali seguono la verifica e la correzione degli errori; queste due fasi richiedono sovente di attivare contatti con alcuni dei soggetti che hanno fornito le informazioni. Al complesso di queste attività si fa usualmente riferimento con il termine *editing*. Per gli istituti che hanno il compito di produrre statistiche per l'esterno, il conseguimento di prefissati standard in termini di qualità del prodotto non deve pregiudicare la tempestività nel rilascio delle informazioni. Quando i dati da elaborare sono numerosi e ad elevata frequenza, conciliare questi obiettivi risulta estremamente difficile e costoso.

L'esigenza di una maggiore efficienza nell'impiego delle risorse, sia da parte dei rilevatori sia dei soggetti rilevati, nel controllo della qualità dei dati ha stimolato presso gli istituti produttori di statistiche lo sviluppo di strategie di *editing* selettivo. L'idea di fondo di queste strategie è quella di fissare gli obiettivi primari di ciascuna rilevazione in termini di accuratezza di alcune statistiche di output. L'utilizzo di una funzione di *scoring* consente di stimare il guadagno di precisione

dell'output ottenibile mediante le attività di controllo e rettifica. Si eseguono prioritariamente i controlli sui dati di input che hanno lo *score* e quindi l'impatto atteso più rilevante sulla statistica finale: in tal modo è possibile concentrare le risorse sulle verifiche essenziali. I guadagni di efficienza ottenibili con tali tecniche sono elevati quando i dati di input sono generati da distribuzioni asimmetriche o quando è fortemente asimmetrico il sistema di pesi associati ai dati elementari nel calcolo delle statistiche finali. La forma funzionale dello *score* dipende dalla statistica finale a cui si vuole indirizzare: nel caso del tasso di variazione temporale di un aggregato A costituito dalla somma di N componenti:

$$\Delta A_t = \frac{\sum_{i=1}^N c_{i,t}}{\sum_{i=1}^N c_{i,t-1}} - 1 \quad (1)$$

lo *score* può essere calcolato utilizzando il termine lineare dello sviluppo in serie di Taylor. Il risultato, per la generica componente i-ma, è espresso dalla formula seguente:

$$S_{i,t} = \frac{\partial(\Delta A_t)}{\partial c_i} = \left( \frac{1}{\sum_{j=1}^N c_{j,t-1}} \right) \left( \Delta c_{i,t} - \Delta c_{i,t-1} \cdot \frac{\sum_{j=1}^N c_{j,t}}{\sum_{j=1}^N c_{j,t-1}} \right) \quad (2)$$

dove:  $\Delta c_{i,t}, \Delta c_{i,t-1}$

rappresentano le differenze tra i valori osservati e i corrispondenti valori attesi (ad esempio sulla base delle previsioni del modello) e individuano le anomalie della componente i-ma per i periodi t e t-1.

Nel caso in cui, durante la fase di controllo dei dati rilevati per il tempo t, i valori del periodo precedente possano essere considerati definitivi, ovvero non affetti da errori, la forma funzionale dello *score* si semplifica, in quanto l'aggregazione dei contributi è di tipo additivo:

$$S'_{i,t} = \frac{\Delta c_{i,t}}{A_{t-1}} \quad (3)$$

Nel paragrafo successivo viene presentato il contesto operativo che ha suggerito la sperimentazione di tecniche di *editing* selettivo per i controlli di qualità degli aggregati relativi alle statistiche monetarie. La tecnica è applicata secondo una logica di tipo *top-down*, che parte dal massimo livello di aggregazione per scendere alle componenti. Tale processo, che qui viene arrestato a livello delle serie elementari componenti un aggregato significativo, potrebbe essere ulteriormente spinto all'analisi dei dati individuali segnalati dalle istituzioni creditizie.

### **3. Il controllo di qualità delle statistiche bancarie**

I dati utilizzati in questo lavoro sono quelli che le banche inviano mensilmente alla Banca d'Italia per la compilazione degli aggregati monetari e creditizi. Per circa 350 serie, le informazioni individuali – relative ai bilanci di circa 800 banche - vengono aggregate e inviate alla BCE che provvede alla predisposizione delle statistiche per l'area dell'euro.

Le segnalazioni trasmesse dagli intermediari vengono sottoposte a un primo controllo di correttezza formale e di coerenza interna. Dopo che questi due passi hanno prodotto risultati soddisfacenti, le informazioni vengono elaborate e sottoposte a verifiche a livello aggregato. Quest'ultimo processo, che spesso richiede la verifica di taluni dati presso le banche segnalanti, deve essere svolto nell'arco di un solo giorno lavorativo. La sperimentazione di tecniche efficienti per il controllo di qualità risulta in questo contesto particolarmente opportuna.

Per una prima verifica è stato scelto il macro aggregato delle contropartite della moneta M3, in quanto esso ha una struttura relativamente semplice e presenta alcune componenti soggette a errori di segnalazione. L'aggregato (cfr. Banca d'Italia, 2000) si compone di 17 serie relative a impieghi bancari al settore detentore di moneta e al portafoglio titoli delle banche (obbligazioni private e azioni). Per ciascuna serie e per ciascuna banca è stato possibile effettuare un confronto tra i dati corrispondenti al primo invio e quelli attualmente presenti nella base dati aziendale, per i quali il processo di validazione è presumibilmente concluso. Tale elaborazione è stata possibile utilizzando agli archivi in cui sono memorizzate le singole rettifiche prodotte nel tempo dagli intermediari. In questo modo è stata realizzata una simulazione che ha consentito di verificare le prestazioni di un processo di *editing* che parte dai primi dati ricevuti da ogni banca, seleziona quelli da sottoporre a verifica, li sostituisce dove necessario e ricalcola gli indicatori finali. Il confronto tra

i dati precedenti e successivi alle rettifiche consente anche di misurare la correlazione tra gli *outliers* individuati dal modello e gli errori riconosciuti dalle banche.

I dati qui analizzati sono relativi all'ultimo semestre del 2002, periodo per il quale si può ritenere che l'attività di revisione dei dati sia stata completata.

#### 4. La strategia proposta

L'applicazione dei modelli per serie storiche al controllo di qualità dei dati è stata considerata per lungo tempo troppo impegnativa in termini di tempi e di risorse necessarie per la stima dei parametri. L'utilizzo su larga scala di tali modelli è stato favorito dalla disponibilità del software TRAMO/SEATS (cfr. Gomez e Maravall, 1996) che provvede all'identificazione automatica e alla stima dei modelli ARIMA con ottime prestazioni in termini di tempi e di affidabilità<sup>2</sup>. Il modulo TERROR ("TRAMO for errors": cfr. Caporello e Maravall, 2003), rappresenta un'applicazione di TRAMO-SEATS all'identificazione dei valori anomali. Il criterio utilizzato è quello dell'errore di previsione standardizzato: per ciascuna serie sono evidenziati come anomali i valori esterni all'intervallo di confidenza basato sulla statistica *t* di Student generata dal rapporto tra l'errore della previsione (valore stimato - valore segnalato) e la sua variabilità media. Una prima applicazione di questo software alle serie mensili bancarie (Farabullini, 2002) ha suggerito una modifica del criterio per l'individuazione dei valori anomali, per tenere conto dell'impatto dei presunti errori individuali sulle serie aggregate oggetto di segnalazione alla BCE. Le tecniche di *editing* selettivo consentono di affrontare questa esigenza con un approccio formalizzato e sistematico, legando in modo diretto i controlli all'impatto atteso sul valore delle statistiche di interesse.

La sperimentazione qui descritta è stata realizzata in ambiente *Speakeasy*, utilizzando le estensioni *Modeleasy+* per l'interfaccia con il software TERROR e gli algoritmi di calcolo supplementari. Su ciascuna delle serie componenti, nella versione che corrisponde al primo invio da parte degli intermediari, il software TERROR è stato applicato per produrre, sulla base di modelli ARIMA, le previsioni a un passo per il periodo *t*. Tale esercizio è stato ripetuto per 7 mesi, al fine di

---

<sup>2</sup> Nella sperimentazione realizzata, su un aggregato composto da 17 serie mensili disponibili su un intervallo di 8 anni, la procedura di stima con un *rolling horizon* di 13 periodi ha richiesto un intervallo di circa 750 millisecondi su un sistema UNIX RISC6000 S85. Tempi dello stesso ordine si ottengono anche su un PC con 512Mbyte di RAM e 1.8GHz di clock rate. Questo risultato rafforza la sensazione che le prestazioni dell'applicazione non rappresentino un vincolo se si considera anche che nella prassi operativa si dovrebbe elaborare una sola previsione *one step ahead*.

poter simulare su altrettanti periodi l'attività di controllo di qualità dell'ultimo dato disponibile di una serie storica. Definita come statistica di interesse la variazione relativa mensile dell'aggregato corrispondente alle 'contropartite della moneta', si è provveduto al calcolo per ciascun periodo e per ciascuna componente della funzione di *scoring*, sfruttando le distanze tra valori osservati e valori previsti dai modelli. Le componenti sono state quindi ordinate in funzione dei valori assoluti dello *score* simulando un processo di verifica dei dati orientato secondo le priorità così individuate. Le prestazioni della funzione di *scoring* sono state esaminate per entrambe le formulazioni (2) e (3), ottenendo risultati simili; il paragrafo che segue commenta l'output della (2).

## 5. Principali risultati e future direzioni di ricerca

L'applicazione della modellistica ARIMA con identificazione automatica del modello ha prodotto risultati molto variabili in termini di accostamento alle serie: gli errori relativi di previsione a un passo, valutati come media, per ciascuna componente, sui periodi considerati, oscillano da un minimo dello 0,70 a un massimo dell' 81 per cento; al netto di questi due valori estremi, l'errore relativo sulle rimanenti quindici serie componenti è pari, in media, al 10 per cento. E' da notare che la capacità previsiva del modello non è importante in termini assoluti, ma relativi, per la capacità di individuare e ordinare le anomalie. I valori previsti dal modello non vengono infatti, di norma, sostituiti con quelli osservati, ma vengono utilizzati come *benchmark* per impostare una graduatoria delle verifiche da effettuare. La variabilità dell'errore di previsione del modello è in parte spiegata dalle discontinuità intervenute nei periodi esaminati su alcune delle 17 componenti<sup>3</sup>. Le corrispondenti serie di 'aggiustamento' sono talvolta disponibili contestualmente al verificarsi del *break*. Nell'esperimento qui descritto, si è preferito valutare la *performance* della procedura nelle condizioni meno favorevoli.

I risultati sono sinteticamente rappresentati nelle figure delle pagg. 11 e 12. Per ciascuna delle tre spezzate, il tracciato corrisponde ai valori che assumerebbe il tasso di variazione mensile delle contropartite della moneta durante un processo di *editing* guidato da una funzione di *scoring*: nell'origine, il tasso di variazione è calcolato con i valori segnalati dagli intermediari con il primo

---

<sup>3</sup> Ad esempio, in questa versione del lavoro non vengono corretti i *breaks* causati da cambiamenti nella popolazione segnalante o da variazioni della classificazione della controparte che comportano cambiamenti nella composizione del settore detentore di moneta.



invio dei dati; nei punti successivi, esso viene ricalcolato sostituendo, per una delle 17 serie componenti alla volta, il valore del primo invio con il valore corretto o con una sua stima. Le tre spezzate corrispondono a tre processi di *editing*, guidati da tre funzioni di *scoring*, tutte calcolate secondo la formula (2) ma con diversi criteri di riferimento per il calcolo del *benchmark*. Nel processo (a) la priorità per il controllo delle componenti è stabilita in funzione delle distanze tra i valori osservati e le previsioni dei modelli ARIMA stimati da TRAMO-SEATS; la curva (b) è costruita simulando un processo di verifica guidato dall'indice di anomalia proposto da TERROR, basato sugli errori standardizzati del modello; il caso (c) ricostruisce ex-post il processo più efficiente, basato sulle rettifiche effettivamente pervenute. I valori sono normalizzati, rispetto al valore definitivo del tasso, che incorpora tutte le rettifiche degli intermediari. Alla fine del processo di verifica, tutti i valori sono pertanto pari a 1. In termini assoluti, l'impatto dei controlli anticipato dallo *score* sulla variazione relativa dell'aggregato oscilla, nel periodo osservato, fra lo 0,1 e il 3 per cento a seconda della componente. Se il modello prevede correttamente, questo divario si può direttamente interpretare in termini di guadagno di efficienza derivante dall'applicazione dell'*editing* selettivo: dando priorità ai controlli a elevato impatto, si stabilizza velocemente l'aggregato sul valore corretto. Una misura sintetica dell'affidabilità della graduatoria basata sui modelli ARIMA è stata ottenuta attraverso il confronto con la sequenza ottimale ricostruita ex-post sulla base delle rettifiche effettivamente segnalate dalle banche. L'indice di cograduazione di Spearman è riportato in ciascun grafico.

Le indicazioni fornite dai grafici e dall'indice mostrano che anticipare l'entità delle revisioni e misurarne l'impatto sulle statistiche di sintesi introduce significativi guadagni di efficienza nelle attività di verifica dei dati. Per tutti i mesi su cui è stata condotta la simulazione, con il processo (a) la velocità di convergenza delle stime ai valori definitivi è infatti superiore a quella del processo (b) guidato dagli errori relativi del modello. L'elevata cograduazione tra il processo guidato dal modello e quello ricostruito ex-post sulla base delle rettifiche effettivamente trasmesse dalle banche fornisce un'ulteriore misura della validità della strategia adottata. Per dati vincoli sulle risorse disponibili per le attività di *editing*, il procedimento selettivo può dunque migliorare la qualità dell'output a parità di numero di controlli, ovvero ridurre il numero dei controlli senza significative perdite informative sulle statistiche finali.

Partendo da questo approccio di base, è possibile costruire funzioni di *scoring* più elaborate e/o estenderne l'applicazione ai dati elementari segnalati dai singoli intermediari. Ad esempio, è possibile introdurre nello *score* una probabilità a priori di errore che può differenziare serie o

intermediari; oppure è possibile combinare *scores* orientati a differenti statistiche di output in uno *score* globale per ottenere un indice che tenga conto delle esigenze di precisione su statistiche di output diverse per contenuto, livello di aggregazione o forma funzionale. Correlata con la probabilità che il valore osservato sia errato è l'entità dell'errore relativo di previsione, che esprime la distanza del dato più recente rispetto all'andamento 'normale' della serie. Sarebbe pertanto utile combinare l'approccio al controllo di qualità basato sullo *score* con quello più tradizionale basato sul test t per la significatività degli errori di previsione. Tale combinazione può tradursi in un indice sintetico del tipo:

$$Werror(t) = |score(t)| + \alpha * |relerr(t)|$$

dove  $relerr(t)$  è la statistica t data dal rapporto tra l'errore di previsione per l'osservazione al tempo t e la deviazione standard degli errori di previsione sull'intera serie storica.

Tale indice sintetico potrebbe essere interpretato come indice di anomalia ponderato con l'impatto sulla statistica finale, o anche come indice di impatto del dato anomalo ponderato con una misura dell'anomalia. Il valore dei pesi da attribuire ai due addendi dovrebbe essere determinato sulla base dell'esperienza e di situazioni contingenti che possono suggerire di alzare la soglia di sensibilità sull'uno o l'altro fattore. In linea di massima, il peso da attribuire al parametro  $relerr(t)$  dovrebbe essere tale da rappresentare una correzione alla graduatoria redatta secondo lo *score* per tener conto di anomalie che, pur essendo di dimensione relativamente piccola, hanno ripercussioni importanti sulla statistica finale nonché sulla qualità delle singole serie componenti.

Per un'analisi in parallelo dei due fattori, si potrebbe comunque prevedere di organizzare l'output della procedura per il controllo di qualità offrendo alcune opzioni in coda all'esecuzione del programma ERROR. In particolare: ottenuto mediante ERROR il raggruppamento dei dati in *normal*, *possible error*, *likely error* sulla base del test t e delle soglie di sensibilità definite dall'utente, le opzioni aggiuntive potrebbero prevedere - in alternativa - la stampa di:

- tutti gli errori 'possibili' o 'probabili' per la prescelta soglia di sensibilità, in ordine decrescente di valore assoluto dello *score*;
- tutti gli errori 'probabili' per la prescelta soglia di sensibilità, in ordine decrescente di valore assoluto dello *score*;
- tutti i dati con un valore assoluto dello *score* > x (es. x=0,01 se lo score si riferisce a un rapporto), indipendentemente dall'entità dell'errore relativo.

Il numero di sospetti errori stampati per il controllo si può in questo modo agevolmente ridurre o aumentare variando la soglia di sensibilità e il grado di anomalia da evidenziare.

Ulteriori verifiche sulle prestazioni della funzione di *scoring* e l'applicazione estensiva della metodologia alle segnalazioni individuali delle banche costituiscono gli obiettivi per il prosieguo della sperimentazione qui descritta.

## Riferimenti bibliografici

Alwan, L.C. e Roberts, H.V. 1988. *Time Series Modeling for Statistical Process Control*, in: Journal of Business and Economic Statistics n. 6, pp. 87-95.

Banca d'Italia 2000. *Supplementi al Bollettino Statistico: note metodologiche e informazioni statistiche*, n. 33.

Barnett V. e Lewis T. 1998. *Outliers in Statistical Data*, Wiley and Sons,.

Caporello, G. e Maravall, A. 2003. *A Tool for Quality Control of Time Series Data – Program TERROR*..

Cheng, B. e Titterington, D.M. 1994. *Neural networks: a review from a statistical perspective*, in: Statistical Science, n.1, pp.1-54.

Delgado, A. e Prat, A. 1997. *Modeling time series using a Hybrid System: Neural Networks and Genetic Algorithm*, in: Bellacicco A., Lauro N. Reti Neurali e Statistica, Franco Angeli.

Farabullini F. 2002. *I modelli ARIMA per la stima e il controllo delle statistiche delle istituzioni finanziarie e monetarie*, in: Bancaria, n. 2/2002, pp. 36-48.

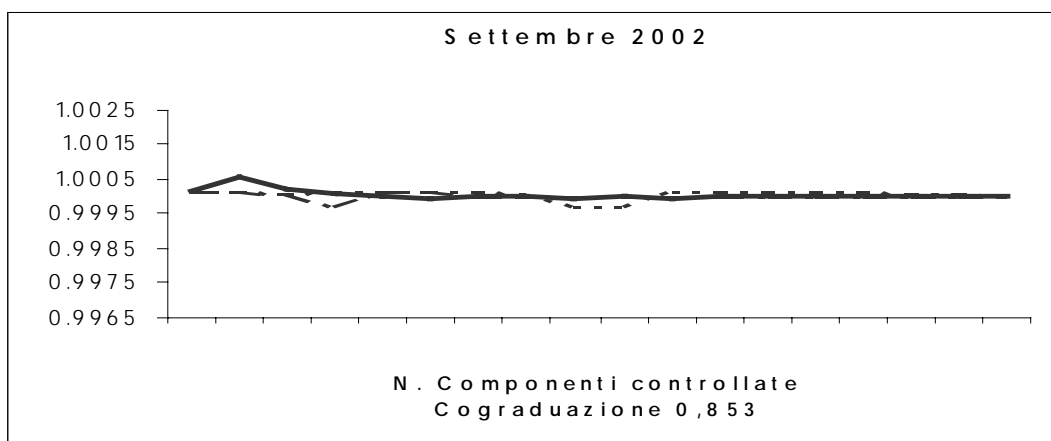
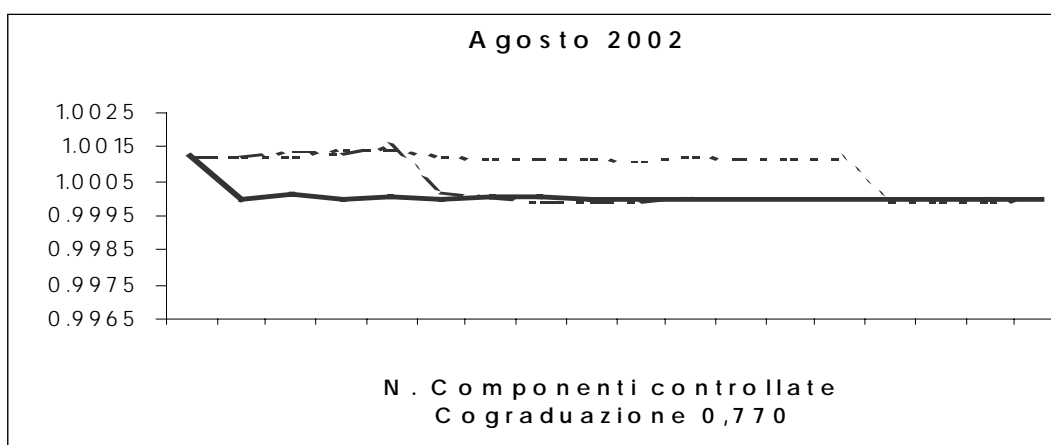
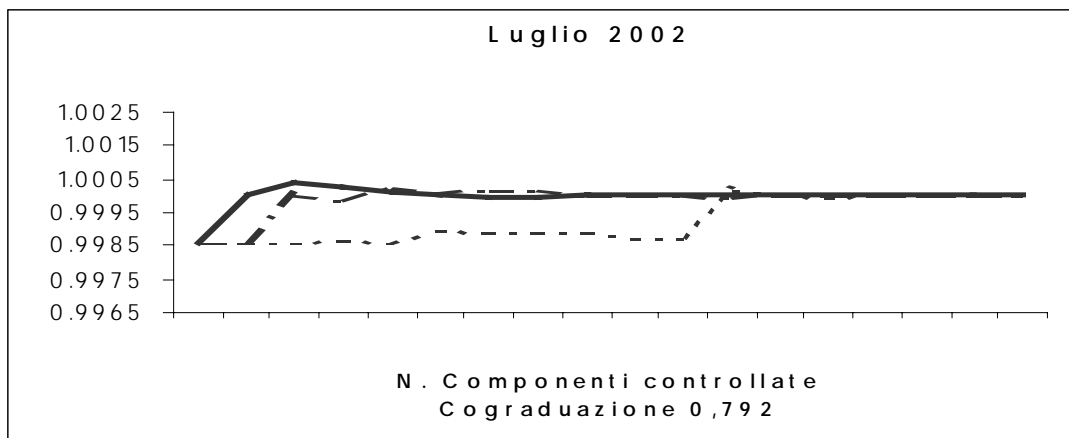
Gomez, V. e Maravall, A (1996). *Programs TRAMO (Time series Regression with Arima noise, Missing observations, and Outliers) and SEATS (Signal Extraction in Arima Time Series). Instructions for the User*. Working Paper 9628, Servicio de Estudios, Banco de Espana.

Landenna, G. e Marasini, D. 1990. *Metodi Statistici Non Parametrici*, Il Mulino, Bologna.

Latouche M. e Berthelot J-M. 1992. *Use of a Score Function to Prioritize and Limit Recontacts in Editing Business Survey*, in: Journal of Official Statistics, vol.8, n. 3.

Lawrence D. e McKenzie R. 2000. *The General Application of Significance Editing*, in: Journal of Official Statistics, vol.16, n. 3,.

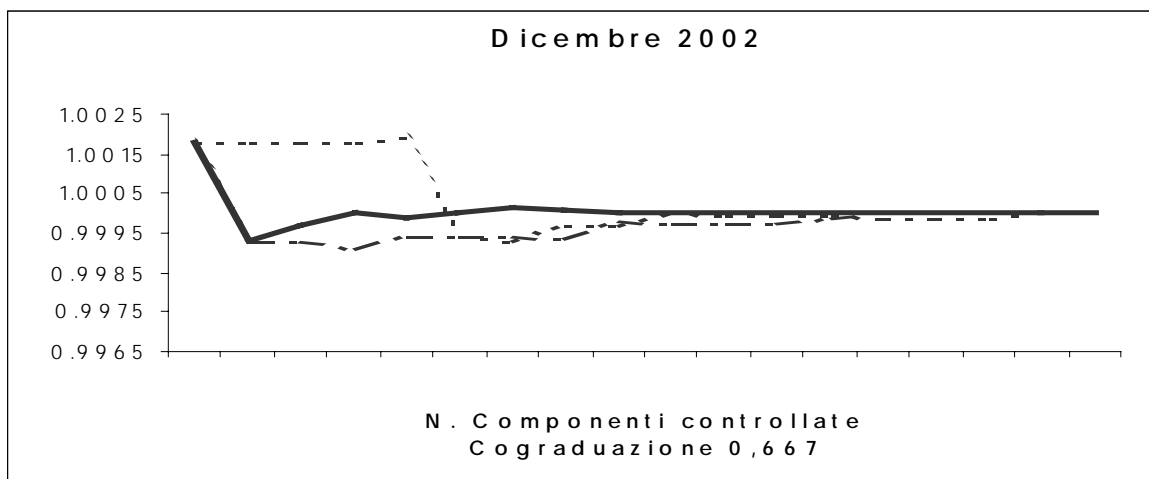
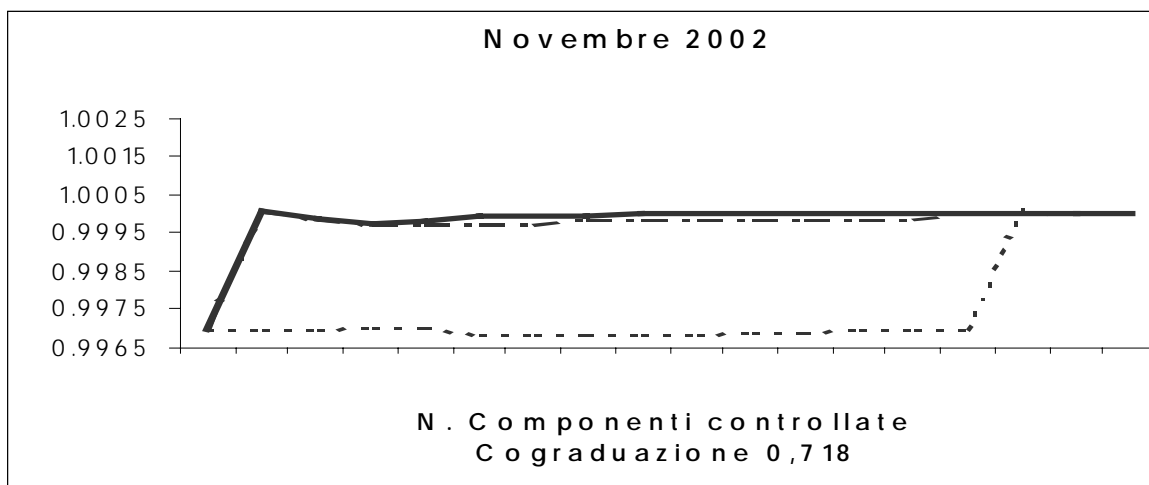
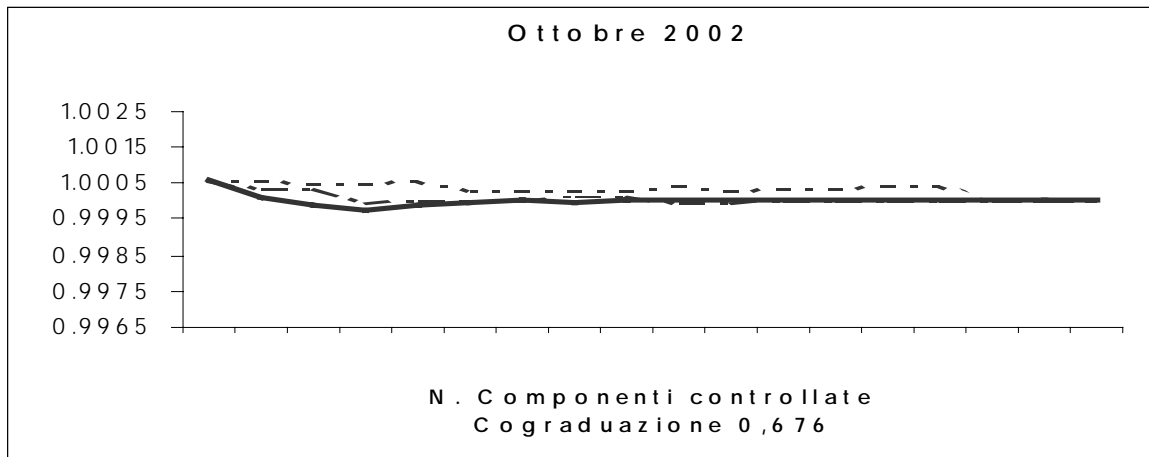
Luzi O. e Pallara A. 2000. *Combining Macroediting and Selective Editing to Detect Influential Observations in Cross-Sectional Survey Data*, in: Statistical Data Editing, ISTAT, n. 6.



**Legenda:**

**Priorità nei controlli secondo:**

- (a) lo *score*
- ..... (b) gli errori di previsione relativi
- (c) le revisioni prodotte dagli enti segnalanti



**Legenda:**

Priorità nei controlli secondo:

- · — · — (a) lo *score*
- ..... (b) gli errori di previsione relativi
- (c) le revisioni prodotte dagli enti segnalanti

*Stampa su carta riciclata*