

Individuazione e trattamento degli outlier mediante il metodo del macroediting

A. Cirianni, S. Coppola, R. Succi

Finalità del macroediting

Il macroediting è finalizzato alla individuazione ed al trattamento dei dati anomali influenti sulla stima del numero indice.

Differenza tra macroediting e microediting

- Il *microediting* è finalizzato al controllo ed alla correzione di tutte le unità elementari;
- Il *macroediting* è finalizzato al controllo ed alla correzione delle sole unità influenti sulla stima del numero indice.

Macroediting

Vantaggi

- risparmio operativo: minore numero di interventi sui dati elementari;
- stessa precisione delle stime rispetto alle tecniche di microediting;
- maggiore tempestività

Svantaggi

manca di garanzia sulla correttezza di tutte le unità elementari

Fasi del macroediting

- Individuazione della variabile obiettivo della rilevazione (variazione tendenziale del fatturato);
- individuazione degli strati e dei domini di interesse;
- definizione della funzioni di controllo macro per la selezione degli strati influenti;
- definizione della funzione di controllo micro per la selezione delle imprese influenti;
- definizione di un criterio empirico per la selezione degli strati mediante la funzione macro;
- definizione delle regioni di accettazione per la funzione di controllo micro.

Funzione di controllo macro

- La funzione di controllo macro (FC1) è il prodotto tra variazione tendenziale del fatturato dello strato (VTFs) ed il peso dello strato nel dominio (Ws).
- La funzione di controllo esprime il contributo di ciascuno strato alla stima del numero indice del dominio.

$$FC_1 = VTF_s * W_s$$

Criterio empirico per la selezione degli strati

- La funzione di controllo macro è finalizzata a selezionare gli strati che permettono di spiegare almeno l'80% della variazione del numero indice del fatturato del dominio.
- La soglia dell'80% è stata scelta come regola empirica.

Funzione di controllo micro: definizione

La funzione di controllo micro (FC_{2i}) è il prodotto tra la variazione tendenziale del fatturato relativa alla singola impresa (VTF_i) ed il peso del fatturato dell'impresa sul fatturato totale dello strato riferito al tempo $t-1$ ($W_{(t-1)i}$)

$$FC_{2i} = VTF_i * W_{(t-1)i}$$

Funzione di controllo micro: finalità

- esprime il contributo di ogni impresa alla stima del numero indice dello strato e permette di individuare le ***imprese influenti*** sulla stima del numero indice.
- la regione di accettazione viene scelta mediante l'analisi grafica (metodo del box plot)
- il processo di individuazione delle imprese influenti si arresta quando le stime si stabilizzano

Individuazione degli outlier influenti

- Le imprese ***influenti*** vengono individuate in base alla funzione di controllo micro.
- Le imprese ***anomale*** vengono individuate mediante il metodo di Hidiroglou-Berthelot (HB).
- Le imprese ***anomale influenti*** sono quelle individuate come outlier dal metodo HB e come influenti dalla funzione di controllo micro.

Individuazione degli outlier mediante il metodo di Hidiroglou-Berthelot

- A. Individuazione della variabile obiettivo (variazione tendenziale del fatturato)
- B. Introduzione di una trasformazione della variabile obiettivo per tenere conto della asimmetria della distribuzione
- C. Definizione dell'intervallo di accettazione (metodo del quartile)

Trattamento degli outlier influenti

- 1) **revisione manuale e contatti telefonici per la correzione degli errori di registrazione o di compilazione;**
- 2) **esclusione degli outlier non rappresentativi;**
- 3) **trattamento degli outlier rappresentativi:**
 - **inclusione;**
 - **esclusione;**
 - **autorappresentazione (le imprese rappresentano solamente loro stesse nella popolazione);**
 - **trattamento outlier come integrazione di MRP**

Tabella di frequenza

| INFLUENTI \ OUTLIER | NO | SI | TOTALE |
|---------------------|---------------|---------------|---------------|
| | NO | 2627 62,2% | 1121 26,5% |
| SI | 129 3,1% | 347 8,2% | 476 11,3% |
| TOTALE | 2756 65,2% | 1468 34,8% | 4224 100% |

Sviluppi futuri

Applicazione di un metodo di macroediting multivariato che considera le variabili fatturato e addetti.