

The problem of outliers in Short Term Surveys on Turnover of Other Services

*Il problema degli outlier nelle Indagini Congiunturali del Fatturato degli
Altri Servizi*

Fernanda Panizon, Alfredo Cirianni

Istituto Nazionale di Statistica (Istat), panizon@istat.it, cirianni@istat.it

Riassunto: Il Regolamento Europeo sulle statistiche congiunturali prevede il calcolo di indicatori trimestrali di fatturato per il settore dei servizi. I dati, raccolti tramite indagine campionaria, devono essere attentamente analizzati per verificare la presenza di eventuali valori anomali che possono pesantemente incidere sugli indici finali: le metodologie utilizzano funzioni punteggio ed il metodo di Hidioglou-Berthelot per mettere in evidenza eventuali outlier, ma la mancanza di veri e propri standard assegna ancora a tale operazione qualche grado di soggettività.

Keywords: short term economic indicators, outliers, macroediting

1. Short term indicators on turnover

Regulation of the European Union requires the member states to produce several short terms economic indicators. For the service sector the main indicators are based on turnover. Different surveys are carried out by Istat collecting data from a sample of firms, about their quarterly revenues of sales and the mean number of employed persons. Based on these data, trend of turnover is calculated year to year as an index, which consider the ratio of turnover in quarter t and quarter $t-4$.

For wholesale trade sector a stratified sample of about 7.000 firms is drawn, which are contacted every quarter for some years. They can answer by mail, fax or web. Mail and fax data are entered by operators while web data are directly stored in our database.

2. Dealing with anomalous values

Data coming from the survey are slightly checked in the preliminary phase of data entry and the main editing phase is done considering macroediting. The most important strata are selected using a (macro) score FC_I to determine if they are influent for the index of their domain.

$$FC_I = VTF_s * W_s \quad (1)$$

FC_I is obtained as the product of trend variation of turnover index number in stratum s (VTF_s) and the weight of turnover in stratum s in its domain (W_s). It can be considered as the contribution of each stratum on the domain estimates of index (I_D).

Then we considered strata that have the higher values of score function (in absolute terms) up to explain at least 80% of trend variation of the domain index..

A second (micro) control function is used, which depends both on the weight of each firms in determining the stratum index and on the trend variation of each firm.

$$FC_{2t} = VTF_i * W_{(t-4)s} \quad (2)$$

FC_{2t} is the product of the trend quarterly variation of turnover for the firm i and the weight of the firm on total turnover of its stratum s at time $t-4$. The sum of FC_2 results in the trend variation index of the reference period t in comparison with the previous $t-4$, that means FC_2 measures the contribution of firm i on the stratum index.

In presence of “high” score function a firm is flagged as influent. The threshold for the score function is determined by means of a box plot graphical method (using SAS INSIGHT) on FC_2

Hidiroglou - Berthelot (H-B) function is then calculate on trend variation to detect anomalous values. If a firm belongs to an influent stratum, if it is influent and is out the H-B threshold it is considered as an outlier and eliminated from the calculation of the index. A manual revision is then done to seek and analyse specific cases of outliers that can be reconsidered, and eventually contacted again (by phone) to confirm data and to obtain some explanation of relevant variation.

Sometimes also a strong trend for a firm can be evaluated as “representative” of other cases in the parent population, and in this case the outlier can be put again in the set of data for the calculation of the index.

3. Conclusions

The whole procedure of macroediting depends on the predetermined parameter C of H-B and the subjective decision of expert revisers that analyse final patterns.

Including or excluding specific anomalous data not considered as outliers, strongly affect final estimation of the index. Some examples are presented to underline the difficulties implied in the editing procedures and the connected risk of error due to the lacking of a standard and tested methodology to deal with anomalous data.

References

- Barcaroli G., Luzi, O., Ceccarelli, C. (1998) *Il macroediting:tecniche di correzione interattiva di variabili quantitative guidata dall'analisi degli aggregati. Il caso del sistema dei conti delle imprese*. Quaderni di ricerca ISTAT, n. 1/1998
- Hidiroglou, M.A., Berthelot, J.M. (1986) Statistical Editing and Imputation for periodic Business Surveys, *Survey Methodology*,12, 73-84
- Lee H. (1995) Outliers in business surveys, in: *Business survey methods*, Cox, Binder, Chinappa, Christianson, Colledge, Kott, John Wiley & Sons, New York, 503-523